English Prosody of Advanced Learners: A Contrastive Interlanguage Analysis

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Philosophie des Fachbereiches 05
der Justus-Liebig-Universität Gießen

vorgelegt von
Karin Puga

aus Kottweiler-Schwanden

2021

Dekan:	Prof. Dr. Thomas Möbius
1.Berichterstatter/in:	Prof. Dr. Joybrato Mukherjee
2.Berichterstatter/in:	Prof. Dr. Sandra Götz
Tag der Disputation:	16.04.2021

Für Anna und Jiří

Table of contents

Acknowledgements

I would first like to express my warmest gratitude to my Ph.D. supervisor, Prof. Dr. Joybrato Mukherjee, to whom I am particularly indebted. Your linguistic expertise was invaluable in formulating the research questions and theoretical constructs. Your critical and honest feedback pushed me to bring my work to a higher level. I benefited from the many presentations I have held during the colloquiums at JLU and the various international conferences, which were only made possible through your generosity and for taking your time for your Ph.D. students.

My special thanks go to Prof. Dr. Sandra Götz, to whom I will always be thankful for her guidance and advice throughout the process of writing this thesis. She has inspired me to become a researcher and she inspired the main idea of the present thesis. I spent many years as her student and later as a research assistant and colleague, and I owe her the basis of my theoretical knowledge, empirical work, and critical thinking in linguistics. I have benefited profoundly from her knowledge, support, and her feedback, which helped immensely to move the thesis forward in a focused manner. She is among the smartest, most knowledgeable and caring persons I know, and I am grateful for having her as my mentor during all stages of Ph.D. work. Thank you, Sandra, for nudging me into the right direction and for helping me to sharpen my focus.

Also, I am very grateful to Rosemary Bock, Robert Fuchs, and Andrea Pešková for their valuable suggestions and discussions on earlier versions of this project.

I have learned a lot from my colleagues and professors of the English department during my time as a Ph.D. student at JLU, especially Christoph Wolk, who was always available for statistical questions. I am grateful to Stefan Gries who reviewed my methodology and results section and who provided me with invaluable feedback and suggestions and who always took his time to discuss statistical questions with me during his time in Giessen. I would also like to thank Kathrin Kircili, Tobias Bernaisch, Bianca Widlitzki, Bridgit Fastrich, Christohph Wolk, Dilini Algama, and Barbara Güldenring for their support and feedback. I would like to thank Barbara Güldenring for proofreading this thesis and for being the most supportive Ph.D. cheerleader.

Additionally, I would also like to thank our six student assistants who helped with the annotation of the data set and for whose critical input has increased the quality of the annotations considerably: Kathrin Kircili, Stefanie Hehner, Karin Yamanaka, Daniela Schulz, Eva-Maria Sarnovski, and Julia Gutschow. The valuable contributions of the annotators were made possible by funding from the TELL-OP project.

My mother has supported me as long as I can remember, pushing me to pursue my dreams and hobbies, whatever they may have been. One of the dreams being to study the English language. Mom, thank you for buying my first Harry Potter book to kick-start my English linguistics career and for your understanding when I was not able to spend as much time with you and the rest of the family, even though I would have liked to. I also want to thank the rest of my family and my friends, my dad, Robert, Marina, Carmen, Shanice, Paula, Martin, Lilli, Jose, and Vanessa for their love and support throughout the Ph.D. work, both emotional and tangible.

List of tables

List of figures

List of abbreviations and acronyms

Aix-MARSEC	(Aix-en-Provence) Machine-Readable Spoken English Corpus
AM	Autosegmental-Metrical
AmE	American English
AmEO	American English Other
AoA	Age of Acquisition
AoAr	Age of Arrival
AoL	Age of Learning
APs	Accentual Phrases
AuSE	Australian English
BrE	British English
BV	Basic Variety
CA	Contrastive Analysis
CAH	Contrastive Analysis Hypothesis
CAT	Communication Accommodation Theory
CDS	Child-Directed Speech
CEFR	Common European Framework of Reference for Languages
cf.	confer
CHECK	checking (pragmatic function)
CIA	Contrastive Interlanguage Analysis
CLT	Central Limit Theorem
CON	continuation (pragmatic function)
CPH	Critical Period Hypothesis
CV	creaky voice
CzE	Czech English
CZ-TOBI	Czech Tone and Break Indices
dB	decibels
DID	difference in distance
e.g.	*exempli gratia*
ed.	editor
eds.	editors
EFL	English as a Foreign Language
ELT	English Language Teaching
EMP	emphasis (pragmatic function)
ENL	English as a Native Language

ERB	Equivalent Rectangular Bandwidth
ESL	English as a Second Language
et al.	*et alii*
etc.	*et cetera*
F0	fundamental frequency
FDS	Foreigner-Directed Speech
FIN	finality (pragmatic function)
FLT	Foreign Language Teaching
GA	General American
GerE	German English
G-TOBI	German Tone and Break Indices
HL	high vs. low tone labels
HOT	handing over of turn (pragmatic function)
HRT	High Rising Terminal
Hz	Hertz
i.e.	*id est*
ibid.	*ibidem*
IL	Interlanguage
IM_PHW	number of ips phw
INSEC	insecurity (pragmatic function)
INTSINT	International Transcription System for INTonation
ip	intermediate phrase
IP	intonation phrase
IP_PHW	number of IPs phw
IPA	International Phonetic Alphabet
IPO	Institute for Perception Research
IPrA	International Prosodic Alphabet
ITAs	International Teacher Assistants
IU	intonation unit
IvIE	Intonational Variation in English
L1	first language
L2	second language
LCL	Learner Corpus Linguistics
LDH	Linear Decline Hypothesis
LeaP	Learning Prosody in a Foreign Language
LFC	*Lingua Franca* Core

LILt	L2 Intonation Learning Theory
LINDSEI	Louvain International Database of Spoken English Interlanguage
LINDSEI-CZ	Czech component of LINDSEI
LINDSEI-GE	German component of LINDSEI
LINDSEI-SP	Spanish component of LINDSEI
LIST	listing (pragmatic function)
LLC	London Lund Corpus
LOCNEC	Louvain Corpus of Native English Conversation
LoR	Length of Residence
LTD	Long Term Distributional
MAE-TOBI	Mainstream American English Tone and Break Indices
MDH	Markedness Differential Hypothesis
MLR	mean length of run
NBrE	Northern British English
NE	Native English
NG	Native German
NN	non-native
NNS	non-native speaker
NS	native speaker
NSV	New South Voices
NWSP	Nationwide Speech Project
PaInTE	Parametric Intonation Events
PAM	Perceptual Assimilation Model
PAM-L2	Perceptual Assimilation Model for non-native speakers
PCA	Principle Component Analysis
PDQ	Pitch Dynamism Quotient
phw	per hundred words
RP	Received Pronunciation
RQ	research question
SAmE	Southern American English
SBrE	Southern British English
SLA	Second Language Acquisition
SLH	Strict Layer Hypothesis
SLM	Speech Learning Model
SoCal	Southern California
SpE	Spanish English

SPL	sound pressure level
spm	syllables per minute
SP-TOBI	Spanish Tone and Break Indices
SSBrE	Southern Standard British English
ST	semitone
TL	target language
ToBI	Tone and Break Indices
TOEFL	Testing of English as a Foreign Language
TONES_PHW	tone frequency phw
UNB	Urban Northern British
wpm	words per minute
WPS	words per second
WPS_IM	words per second per ip
WPS_IP	words per second per IP
YOEAS	years of English at school
YOEAU	years of English at university

List of prosodic symbols (Tone and Break Indices (ToBI))

-	phrase accents
%	boundary tones
*	pitch accents
+	combination of bitonal pitch accents
H	high tone
L	low tone

Abstract

The present study investigated the production of prosody of 135 learners of English with the aim of identifying the acoustic properties of f0 in advanced Czech, German, and Spanish Learner English (n=45 each). Their performance was compared to that of monolingual native speakers of British (n=45) and American English (n=45). The main focus of the study concerns prosodic features of declarative utterances elicited from the second language (L2) and native spontaneous English speech (dialogic and monologic), as well as the influence of interviewer questions on some of the declarative utterances. Thus, it is essential to examine entrainment at the phonetic level within spontaneous interactions between interlocutors from the different first languages (L1s) and dialect backgrounds in the case of the native speakers. Within the framework of the autosegmental-metrical approach (Beckman and Pierrehumbert 1986), production of prosody was examined in different tonal events: pitch accents, boundary tones, and their tone heights (f0 measures). In order to achieve a phonetic and phonological description of these tonal events, it is important to investigate the domain of intonational phrasing (i.e. intonation phrases (IPs) and intermediate phrases (ips)) and the patterning of these tones into tunes. Additionally, the study seeks to clarify the relationship between prosodic form and socio-pragmatic meaning, necessitating the analysis of form-function correlations. The present study is one of the first[1] to discuss the findings within a theoretical framework specifically designed for intonation, i.e. the *L2 Intonational Learning Theory* (LILt, cf. Mennen 2015).

The findings are discussed in light of all extralinguistic variables known about all native and non-native speakers (n=225 in total), while the different interlanguages are compared to and with each other and to the two native varieties of English, thus, applying a *Contrastive Interlanguage Analysis* (Granger 1996, 2015). Results reveal a marked contrast between native and non-native speech production of nearly all prosodic events. Compared to the native speakers of English, most of the L2 speakers tend to produce longer ips in seconds, a slower speech rate within intonation units (IUs), more high edge tones, a narrower f0 span for high-low tunes, and a wider and higher f0 for high edge tunes. The only native-like prosodic features are tone frequency, the f0 level of high-low tunes, and IU frequency (for Spanish English learners only). In order to better explain these results, a multifactorial regression analysis of learning and recording context variables was performed, e.g. speaking style, age, gender, stays abroad, pragmatic functions (continuation, listing, etc.), speech acts (statements vs. answers), etc.. Main effects that explain the variation in L2 prosody are found in speaker gender and certain fluency variables, while minor effects were found for proficiency level, age, speaking style, and stay abroad. The results seem to support some predictions made by the LILt model (Mennen 2015). While the learners were found to differ on nearly all intonational dimensions of the LILt model and all prosodic variables investigated, the analyses of individual differences by learner-variety revealed specific trends. For instance, while the Spanish English speakers deviate mostly on the realizational level, i.e. speech rate per IU and f0 range in the different tunes, Czech and German English learners differ the most on the distributional and functional level with the choice of tonal categories (same trend), the variety of pragmatic functions chosen for these tonal categories (similar

[1] See also Ward and Gallardo (2017).

trends), and IU frequency (opposite trend). Based on these results, mostly proficiency-related explanations were given rather than L1-based influences, since all learners exhibited similarly deviant prosody.

In sum, this study presents a quantitative mixed-method analysis of f0 in spontaneous native and non-native speech by identifying patterns of variation and universals in the phonological and phonetic form of f0, as conditioned by extralinguistic and contextual factors. Such an analysis informs our understanding of prosodic variation in native and non-native speech, as well as which factors can predict such variation. The study has implications for theoretical and applied linguistics. The results of the present study can fruitfully inform areas of relevance, such as theoretical second language acquisition processes, language assessment, language teaching, development of computer-based language teaching materials, and speech technology.

Abstract - German

Die vorliegende Studie untersuchte die prosodischen Muster von 135 Englischlernern mit dem Ziel, die akustischen Eigenschaften der Grundfrequenz in fortgeschrittenem tschechischem, deutschem und spanischem Englisch (jeweils n=45) zu identifizieren. Ihre Performanz wurde mit der von einsprachigen Muttersprachlern des Britischen (n=45) und Amerikanischen (n=45) verglichen. Dabei lag der Fokus der Studie auf prosodischen Merkmalen deklarativer Äußerungen, welche aus der spontanen Produktion der Zweitsprache (L2) bzw., im Falle der Kontrollgruppen, der Muttersprache hervorgehen, sowie auf dem Einfluss von Interviewerfragen auf einige der deklarativen Äußerungen. Daher ist es wichtig, die Annäherung der Grundfrequenz (*prosodic entrainment*) auf phonetischer Ebene innerhalb spontaner Interaktionen zwischen Gesprächspartnern aus den verschiedenen Muttersprachen (L1s) und dem Dialekthintergrund bei den Muttersprachlern zu untersuchen. Im Rahmen des autosegmental-metrischen Ansatzes (Beckman und Pierrehumbert 1986) wurde die Produktion von Prosodie in verschiedenen Tonereignissen untersucht: Tonhöhenakzente, Phrasen- und Grenztöne und deren Tonhöhen (f0 Maßnahmen). Um eine phonetische und phonologische Beschreibung dieser Phänomene zu erhalten, war eine Untersuchung des Bereichs der Intonationsphrasierung (Intonationsphrasen (IPs) und Intermediärphrasen (ips)) und der Strukturierung dieser Töne zu Sprachmelodien erforderlich. Um schließlich auch die Beziehung zwischen prosodischer Form und sozio-pragmatischer Bedeutung bestimmen zu können, wurde auch eine Analyse der Form-Funktions-Korrelationen durchgeführt. Die vorliegende Studie ist eine der ersten, die die Ergebnisse innerhalb eines theoretischen Rahmens diskutiert, der speziell für die Intonation entwickelt wurde, nämlich der *L2 Intonational Learning Theory* (LILt, vgl. Mennen 2015).

Die Ergebnisse werden vor dem Hintergrund aller extralinguistischen Variablen diskutiert, die über die Muttersprachler und Nicht-Muttersprachler (insgesamt n=225) bekannt sind, während die verschiedenen Lernervarietäten im Rahmen einer kontrastiven Interimsprachenanalyse (Granger 1996, 2015) miteinander und mit den beiden Muttersprachlergruppen des Englischen verglichen werden. Die Analyse ergibt bei fast allen prosodischen Ereignissen einen deutlichen Kontrast zwischen muttersprachlicher und nicht-muttersprachlicher Sprachproduktion. Im Vergleich zu den englischen Muttersprachlern erzeugen die meisten L2-Sprecher tendenziell längere ips in Sekunden, eine langsamere Sprechgeschwindigkeit innerhalb von Intonationseinheiten (IUs), vermehrte Phrasen- und Grenztöne, eine engere f0-Spanne für *High-Low*-Melodien und eine breitere und höhere f0 für Phrasen- und Grenztöne. Die einzigen muttersprachlich-ähnlichen prosodischen Merkmale der Lerner sind die Tonfrequenz, das f0-Level von *High-Low*-Melodien und die IU-Frequenz, wobei Letzteres lediglich auf spanische Lerner des Englischen zutrifft. Um die Ergebnisse besser erklären zu können, wurde eine multifaktorielle Regressionsanalyse durchgeführt, die folgende Faktoren beinhaltet: Sprechstil, Alter, Geschlecht, Auslandsaufenthalte, pragmatische Funktionen (Fortsetzung/Kontinuität, Auflistung usw.), Sprechakte (Aussagen vs. Antworten) usw. Die Haupteffekte, die die Variation der L2-Prosodie begründen, liegen im Geschlecht der Sprecher und in bestimmten Flüssigkeitsvariablen, während dem GER-Niveau, dem Alter sowie dem Sprechstil und einem möglichen Auslandsaufenthalt lediglich ein geringfügiger Einfluss attestiert werden kann. Insgesamt wurden einige der Annahmen des LILt-Modells (Mennen 2015), die auf Grundlage des vorliegenden Datensatzes testbar waren,

durch die Korpusanalyse untermauert. Während sich die Lerner in nahezu allen Intonationsdimensionen des LILt-Modells und in allen untersuchten prosodischen Variablen unterschieden, ergaben die Analysen klare gruppenspezifische Trends: Während die spanischen Lerner am deutlichsten auf der Realisierungsebene von der Muttersprachlernorm abweichen (d.h. in Bezug auf die Sprechgeschwindigkeit pro IU und den Grundfrequenzumfang in den verschiedenen Melodien), weichen die tschechischen und deutschen Lerner am stärksten in der Verteilungs- und Funktionsebene durch die Auswahl der Tonkategorien ab, wobei für beide Gruppen der gleiche Trend beobachtet werden konnte. Auch die Abweichungen bezüglich der Vielfalt der für diese Tonkategorien ausgewählten pragmatischen Funktionen sowie die Häufigkeit der IUs zeigen entscheidende Unterschiede zu den muttersprachlichen Ergebnissen, wobei zwischen den beiden Lernergruppen ähnliche Trends bzw. – bezüglich der Frequenz der IUs – ein entgegengesetztes Verhalten festgestellt werden konnte. Basierend auf diesen Ergebnissen wurden eher kompetenzbezogene (anstelle von L1-basierten) Erklärungen gegeben, da alle Lerner eine ähnlich abweichende Prosodie etabliert zu haben scheinen.

Zusammenfassend präsentiert diese Studie eine quantitative Analyse der Grundfrequenz mit gemischten und multifaktoriellen Methoden. Dabei werden die phonologische und phonetische Form der Grundfrequenz analysiert, um Variationen sowie Universalien im muttersprachlichen und interimssprachlichen spontanen Sprachgebrauch aufzudecken, die zusätzlich durch extralinguistische und kontextuelle Faktoren bedingt sind. Eine solche Analyse trägt zum Verständnis der prosodischen Variation in der muttersprachlichen und nicht-muttersprachlichen Sprache bei und deckt überdies auf, welche Faktoren eine solche Variation vorhersagen können. Die Studie hat somit Auswirkungen auf die theoretische und angewandte Linguistik, da die Ergebnisse für Bereiche wie theoretische Zweitspracherwerbsprozesse, Sprachbewertung, Sprachunterricht, Entwicklung computergestützter Sprachunterrichtsmaterialien und Sprachtechnologie relevant sind.

1. English prosody: An overview

1.1 Introduction

When we speak in our native language (L1), we mostly do not pay attention to the form of our speech (i.e. which words we choose), but rather place emphasis on *how we use it* (i.e. the tone we use) to get the meaning of our message across to our interlocutors. By doing so, we subconsciously make use of prosodic cues to support the meaning of our message. While some individuals would be described by laymen as having a monotonous speaking style even in their L1, others are described as having a more 'lively intonation'. Not only is there considerable interspeaker variability, but there are also regional differences (e.g. Armstrong et al. 2015; Tench 2015; Fuchs 2018), as well as several regional sub-varieties, which can differ in terms of their prosodic features (e.g. Clopper and Smiljanic 2011). Additionally, prosody is considered highly context-dependent and differs from speaking style to speaking style (e.g. Hirschberg 2000; Mukherjee 2001; Lowry 2002; Grabe 2004; Gut 2009, 2017). Considering that considerable differences between individual native speakers (NS) and native varieties of English exist, how can we investigate non-native speakers (NNS) homogeneously? The present study attempts to answer this question, among others, and, in doing so, compares two native varieties of English, i.e. British English (BrE) and American English (AmE) with each other, as well as describes how different learner groups (L1 Czech, L1 German, and L1 Spanish) approximate to these target varieties of English. In order to find stylistic differences, the study investigates learners' production of prosodic patterns in declarative utterances in two discourse types (monologues and dialogues) in spontaneous speech. Different extralinguistic factors are considered in the analysis along with the variety of English the learners aim at approximating to,[2] in order to find correlations between the prosodic profiles of the learners and the respective native variety.

Advanced learners of English will likely have a vested interest in improving on prosodic features, which is part of the attempt to perfect their English language skills. This would only include a quite small proportion of the entire learner population, since preparing students for "successful communication" outside the classroom (Kennedy and Trofimovich 2008: 460) is the main goal of most *English as a Foreign Language* (EFL) teachers. Prosody is mostly neglected in language teaching, even though the interest and implementation of prosody teaching is growing (cf. "intonation considered as "luxury" in language teaching" Chun 1998: 74; see also Levis 1999). While L1 transfer phenomena in the second language (L2), leading to a perceived "foreign accent", may not be as vital for successful communication, they could, however, in the worst case, offend NSs. A foreign accent may be signaling intonation patterns with different emotional states and attitudes, or might be interpreted as impolite both by native and non-native listener (cf. Kasper 1981; Loveday 1982: 93; Wells 2014: 133; Ward and Gallardo 2017: 21). Therefore, the present study mainly focuses on these areas of prosody in which even highly advanced learners of English deviate from two different NS groups. It is necessary for language teachers to recognize that prosody is important for 'successful communication' and not just for sounding native-like, since

[2] For some this might be a conscious process, while for other learners it might be more unconscious. This information, however, is unknown in the present data set unless the learner specifically mentions it in their recordings.

misunderstandings can arise from transferring L1 prosodic patterns to the L2. The main reasons to include prosody in *Foreign Language Teaching* (FLT) are to reduce an L2 accent and to increase comprehensibility as well as fluency (cf. Derwing and Rossiter 2003). In early childhood and throughout our lives we become attuned to speech regularities, such as the sound and prosody of our L1 (cf. Ortega-Llebaria and Colantoni 2014: 332). Not only are NSs able to discriminate native sounds quickly, but they produce them with ease as well. When non-native (NN) sounds are heard, however, NSs might have processing difficulties even though they might not be able to exactly determine what makes the learner sound 'foreign' (cf. Taylor 1993: 2). Acoustic phenomena remain as pre/categorical percepts in our consciousness for no more than a fraction of a second (Kallman and Massaro 1983: 322) and our decoding mechanisms are geared primarily to the extraction of communicatively relevant information (the semantics of an utterance, its significance for the ongoing communication act) (cf. Barry 2007: 97). Especially since prosody does not have an orthographic transcription in English, it is even more difficult to describe deviances in prosody (cf. Derwing and Rossiter 2003). The way we speak is a kind of "audible business card" (Hirschfeld and Trouvain 2007: 171) and our individual way of speaking reflects our personality and attitude. Prosodic deviances shape the attitudes towards speakers. For instance, the relatively flat and low intonation of German learners of English may make them sound "bleak, dogmatic or pedantic, and as a result, English listeners may consider them uncompromising and self-opinionated" (Trim 1988, as quoted in Grabe 1998). Most importantly, prosodic deviances in non-native speech can contribute to a perceived foreign accent, ill-formed stereotypes, and/or even impede communication, intelligibility, and comprehensibility (Derwing and Munro 1997; Laures and Weismer 1999; Jilka 2000, 2007; Derwing and Rossiter 2003; Trofimovich and Baker 2007; Mennen et al. 2014; Tench 2015: 11; Moyer 2018: 5). Especially prosodic errors seem to substantially contribute to a loss of intelligibility, more so than segmental errors (e.g. Anderson-Hsieh et al. 1992; Munro and Derwing 1999). However, in the context of the present study, intelligibility and comprehensibility will be taken for granted, considering that the learners included in the study have been selected on the basis of being advanced learners of English. Therefore, the focus will not be on prosodic cues that impede the comprehensibility or intelligibility of their speech, but rather describe what exactly it is that creates the immediate impression of 'foreign-soundingness' (cf. Granger 1996: 43; Callies and Szczesniak 2008) in non-native English speech production in terms of their prosody. Therefore, the present study mainly analyzes prosodic variables concerned with production of non-native and native prosody. Non-verbal prosodic variables are not part of the analysis.

The present study is divided into six sections. Section 1 presents the aims and research questions, defines and discusses the most relevant linguistic terminology, discusses prosody in *English Language Teaching* (ELT) and prosody-related *Second Language Acquisition* (SLA) theories and models, and several extralinguistic factors that influence prosody production are reviewed.

The second section includes the theoretical basis and a review of previous research on selected features of native and non-native prosody, i.e. tones and tunes, intonational phrasing, f0 range, and uptalk tunes. First, models of intonational structure and theoretical frameworks of prosody are discussed, and the main intonational model (autosegmental-metrical approach) this study is based on is described in more detail. Following that, the native languages of the three learner groups Czech, German, and Spanish are compared to English prosody in general. Then the general problem areas of L2 prosody production of Czech, German, and Spanish learners are presented and previous research on these learner groups is reviewed and evaluated. This is followed by a review of research on the different selected aspects of the form and function of prosody, in which native and non-native English are compared in four sections (Sections 2.5-2.8). Each section is structured in the same way, i.e. first the prosodic feature is described in native English (NE), the following sub-chapter describes the feature in native Czech, German, and Spanish as compared to English, and finally the prosodic feature is described in non-native speech. Furthermore, each sub-chapter situates the prosodic feature in context, i.e. regional, social, and stylistic variation, entrainment, the effects of para- and extralinguistic variables, and the acquisition process of the specific prosodic variable. All of these aspects will be discussed in the interpretation of the findings in Section 5.

Section 3 gives an overview of the corpus data, the experimental set-up, annotation procedure, as well as the linguistic and extralinguistic variables the study analyzes. Each of the four main prosodic variables is defined and it is explained in detail how these features were operationalized. As the title of the present study already foreshadows, it can best be described as a mixed-methods (corpus-based and explorative), multifactorial, and quantitative analysis of non-native and native prosodic features. Finally, the section concludes with the conditions and the limitations of the present study.

Section 4 summarizes the main findings of the contrastive (interlanguage) analysis of the productive prosody variables and revisits the research questions. The individual variables, such as intonational phrasing, f0 range, uptalk tunes, pitch accent, edge tones, and their pragmatic functions, will be presented and compared to each other within the native and non-native speaker populations. Deviations and approximations to the target norm will be considered when discussing the non-native data. The aim of the present study is to identify, describe, and explain prosodic features of the spoken interlanguage of advanced Czech, German, and Spanish learners of English which may make them sound ‘foreign’ to NSs of English, based on the quantitative production analysis. In a second step these results will be compared to an analysis of AmE and BrE prosodic features.

Section 5 summarizes and discusses the findings of this study, discusses relevant SLA predictors and theories, presents a multifactorial model of non-native prosody that explains the results, as well as some language-pedagogical implications derived from the findings are given.

A brief conclusion of the results is given in Section 6 and the broader implications of the findings of the present thesis are also put forward. Finally, this section gives an outlook for future research on native and non-native prosody.

1.1.1 Aims and research questions of the current research

The international interest in L2 prosody has grown over the last two decades, despite still being demonstrably behind the progress that segmental studies of non-native speech have made. Prosodic research has been plagued (and to some extent still is) by fundamental issues involving which framework to choose for prosodic analysis, which phenomena to analyze, and which methods to use (cf. Ladd 2008: 3). While many of these issues still persist, some issues have been largely resolved in the past decade. For instance, a positive change is represented by the increased use of the *Tone and Break Indices* (ToBI) annotation system (Silverman et al. 1992), which is a marked improvement over the previous prevalence of many different annotations systems for intonation, which made comparisons between studies difficult. Linguistic research on prosody in the past was quite limited in the prosodic features analyzed, research methods used, and its application to language teaching and speech technology. Only until fairly recently have quantitative corpus-based approaches to the analysis of prosodic features been adopted in the study of L2 prosody (e.g. Ramírez Verdugo 2002, 2003, 2005, 2006a, b; Gut 2009). Nevertheless, a more detailed and fully quantified view on L2 prosody is still missing. Therefore, the present study attempts to add to this growing body of research and to fill this research gap by using quantitative corpus-based approaches in describing and interpreting the form, function, and distribution of four different prosodic parameters (tones/tunes, intonational phrasing, f0[3] range, and uptalk). In doing so, the present study takes a phonetic, phonological, pragmatic, as well as sociolinguistic approach to analyzing native and non-native English speech. The specific aims are as follows:

1) To provide a systematic overview of prosody in non-native and native speech
2) To show prosodic deviances in non-native speech and among natives
3) To describe patterns of intonational phrasing and use of f0 range
4) To illustrate the tone inventory (edge tones and pitch accents), combinations of tones (tunes), and their pragmatic meaning in a given context
5) To explore native and non-native prosody in its context (discourse, speaking style, paralinguistic features, other linguistic features, etc.)
6) To develop a multifactorial model of non-native prosody that explains the results

The last aim (6) points toward one of the primary goals of theory building in intonational phonology as well as L2 prosody and SLA theory. The main research questions of the present study are the following:

1) What are the structural and functional features of prosody in the spoken interlanguage of advanced learners of English with different L1 backgrounds?
2) Can universal features be observed across different language families?
3) a) To which extent do these learners diverge from the native speakers' prosodic patterns or adopt language-appropriate values in spontaneous speech?
   b) If the learners deviate from native productions, what are possible reasons? What is the sociolinguistic distribution of these features?
4) What makes the prosodic features of foreign language learners so foreign-sounding?

The answers to these questions will add to the growing body of research in L2 prosody. Throughout the study, these research questions will be more fine-grained and defined in the methodology. The present study is a multidimensional study of f0 in its phonetic as well as phonological representation and realization in different contexts, speaking styles, and individual speakers (i.e. native vs. non-native, male vs. female, regional dialects, etc.). F0 is examined from different perspectives by investigating the domains in which it occurs (IUs),

[3] F0 refers to fundamental frequency.

the phonological representations (ToBI labels) it can take, as well as phonetic implementations (Hertz (Hz), different f0 measures) chosen by speakers, and its occurrences (i.e. position: early influences and entrainment in dialogic speech and final rises in different pragmatic contexts, and quoted vs. non-quoted speech). The main contribution of this thesis is a comprehensive view of f0 in various dimensions and contexts of use. This study is unique for multiple reasons: 1) it compares many different learner languages from three different language backgrounds and two different NS groups; 2) it combines fluency measures with prosodic features; 3) it includes many more linguistic as well as independent speaker and interview variables than previous studies, along with a larger quantity of speakers of 45 speakers per group (five groups in total/including seven different varieties of English). The main goal of the present study is to determine which aspects of prosody characterize intermediate to advanced learners of English from different language backgrounds, as compared to American and British NSs of English. The present study has a theoretical as well as practical relevance. The findings can inform areas such as language teaching and teacher education, classroom methodology, language assessment, material design for FLT, SLA research, learner corpus linguistics (LCL), and speech technology. Additionally, L2 speech learning models in terms of prosody are reviewed and based on the results a multifactorial model of L2 prosody is proposed.

1.1.2 Defining prosody

Every speaker makes use of several prosodic cues during speech production: E.g. duration, pauses, and changes in f0. Through prosody speakers are able to convey linguistic (e.g. utterance type, information structure) as well as paralinguistic (e.g. emotional state) information. There are narrower (linguistically relevant features only) and broader (linguistic and paralinguistic) definitions of prosody depending on whether only linguistically relevant features are regarded as contributing to prosodic forms and functions (e.g. Grice and Baumann 2007; Gut 2009). The present study adopts the broader definition of prosody and includes some paralinguistic features (pause, voice quality, and final syllable lengthening). These features are interesting because languages differ in how they use paralinguistic features of f0 to highlight and segment certain words or phrases (cf. Grice and Baumann 2007: 26). The term *prosody* can be defined as the suprasegmentals of speech, which include the intonation, stress, and rhythm of an utterance. Eskenazi (1999: 64) calls intonation the "glue" that holds together messages. Without intonation speech would become quite unintelligible, since it fulfills various functions, i.e. signaling different attitudes or emotions, resolving syntactic ambiguity, focusing or highlighting, indicating information status (given vs. new) of an utterance, and signaling turn transitions (different pragmatic meanings: continuation and completion). In intonation studies terms like *fundamental frequency* (f0), *pitch*, *intonation*, *prosody*, *suprasegmentals*, etc. are sometimes used interchangeably to refer to the same concept. Especially prosody is frequently equated with suprasegmentals. In this thesis, however, I make a distinction between these terms. *Prosody* will be used as an umbrella term for other interacting phenomena, such as rhythm, f0 range, focus/prominence, and intonational phrasing. The term *suprasegmentals*, is seen as a synonym of *prosody*; however, *prosody* is always the preferred term in the present study, in order to avoid the layering metaphor, which is implied in the term *suprasegmentals* (cf. Beckman and Venditti 2011;

Arvaniti 2017). Prosody is, therefore, not seen as an additional supplement to segmental features, but as an important part of spoken discourse, which is constantly “anchored” with segments (e.g. Ladd et al. 1999). Prosody cannot be separated from segments and is not ‘optional’ or ‘peripheral’. The term *intonation*, on the other hand, will not be regarded as a synonym of *prosody*. In this study, intonation includes *intonational phrasing* and *f0* (perceived, phonological, contrastive, and linguistic). Intonation is, thus, one part of prosody but does not show the entire array of what can be achieved prosodically with a human voice. Intonation is treated as only one part of prosody and it mainly “refers to the linguistically structured and pragmatically meaningful modulation of f0” (Arvaniti 2017: 2). *Fundamental frequency* (f0) is seen as the phonetic correlate to pitch, which can be used for purposes other than intonation (i.e. linguistically structured modulation of f0). While f0 can also be used for the signaling of paralinguistic information (e.g. excitement, nervousness, boredom, etc.), it can structure discourse and signal pragmatic meaning, as well as gender identity and social relationships. Each of the four prosodic features (tones/tunes, intonational phrasing, f0 range, and uptalk) investigated in this study will be defined and discussed on the basis of previous research in Section 2.

1.2 Prosody and English Language Learning in Europe (Germany, Spain, and the Czech Republic)

There are approximately around 1.5 billion NNSs of English in the world (Crystal 2003a: 6). “Given these high numbers, we therefore consider analysis of their language and speech relevant to our understanding of the human capacity for language and speech more generally” (De Leeuw and Celata 2019: 88). As outlined above, the present thesis is concerned with the English interlanguage as spoken by the following three learner groups: Czech, German, and Spanish learners of English. For the remainder of the study the three learner groups and their respective English interlanguages will be referred to as Czech English (CzE), German English (GerE), Spanish English (SpE), Madrid English (MadE), and Murcian English (MurE) to further distinguish the SpE group. Whenever SpE will be used, it refers to both MadE and MurE varieties of English. An explanation for the selection of these three learner groups is required, including description of the circumstances in which English is learned in these countries. Previous studies have tried to explain universals in non-native English prosody, doing so by including different learner populations in their analyses (e.g. Gut 2009, with mixed L1 backgrounds; see also Andreeva et al. 2014a, 2014b for a similar approach). The present study adopts this approach by including learner populations from different language family backgrounds, i.e. Czech (a Slavic language), German (a Germanic language), and Spanish (a Romance language), in order to analyze similarities and differences and to uncover universals in their L2 prosody production. The choice of these three languages was motivated by the fact that I speak these three languages to different degrees of proficiency, which aided the analysis of the prosodic features of these three interlanguages and helped to identify non-native tunes that might originate from the respective L1s.

Even though there are some differences between the education systems of the Czech Republic, Germany, and Spain, the English spoken in these three countries can be categorized in the “expanding circle” of Kachru’s (1996) “Three Circles Model”. According to Kachru’s

(1996) model, English has no official status in these countries and is mainly used in certain institutionalized settings such as schools, universities, or internationally-oriented companies. Other than at school or maybe even at kindergarten (i.e. bilingual kindergartens), children and adults in these three countries are usually only exposed to English if they deliberately seek it out themselves (i.e. if they read English texts or watch English series/movies in their free time) or if they are/were taught English by parents or other relatives, or have access to English-speaking people in their immediate surroundings. Through the internet and other new media, children in these three countries are mainly exposed to a quite limited amount of English through movies, TV series, music, etc. For the majority of the population, therefore, English is learned as one of the mandatory school subjects by adolescents and, in adulthood, is a frequent requirement for a job. Therefore, in Kachru's (1996) terminology, the English spoken by the three learner groups under scrutiny are described as EFL varieties.

The corpus data of the learners of English used in the present study originates from different years, i.e. 2012-2014 for the Czech data, 2004 for the German data, and 2000-2002 for the Spanish data. Even though the Czech learner data was collected a decade after the German and Spanish learner data, no fundamental changes in schooling had occurred, in particular for the treatment of prosody in these EFL classrooms. At the time of corpus compilation, English teaching at school started in grade five in Germany (cf. Brand and Kämmerer 2006). In primary and secondary education in Spain, both English and Spanish is used (Gilquin et al. 2010: 55). At tertiary level in Murcia, instruction is in English, except for some specialist classes (ibid.). During the first educational stage (grades 1-5) in the Czech Republic, English is mostly taught as the first foreign language by one teacher who teaches all subjects. The second stage (6-9) sees the continuation of English foreign language instruction to varying degrees, i.e. students can choose between the four-year elementary school, eight-year grammar school or six-year bilingual grammar school. The six-year bilingual grammar school is naturally the most intense choice of foreign language instruction with about ten foreign language lessons per week and, during the grades 10-13, all subjects are taught in the foreign language. In the Czech Republic students are obliged to learn a foreign language starting from grade three in elementary school and English is the default option, however it is not mandatory and other languages can be chosen (National Institute for Education 2017: 143). Also, there are some schools in the Czech Republic that offer English starting from the first grade (Jakšič and Šturm 2017: 358-359), which explains why some of the Czech learners in the data of the present study have up to 16 years of English in school. In all three countries, English is mainly perceived as being important for work and travel (e.g. Ibarraran et al. 2008).

As laid out above, the EFL classroom is the major source of exposure to the foreign language in all three countries. In the Czech Republic, Germany, and Spain, the students are mainly taught by NNSs of English. 74.19% of EFL teachers in Spain and 95.87% in Germany are L2 users of English (Henderson et al. 2012: 9). When it comes to teaching pronunciation in EFL classrooms in Europe (present and past), teachers often tend to correct (if at all) only the most salient speech errors that students commit while learning an L2 (cf. Euler 2014). These features mostly include the sound system of the English language, such as the pronunciation of <th> and the <r> or lexical stress. However, as mentioned in the introduction, prosody might have an even stronger effect on the perception of a foreign accent than the pronunciation of the individual sounds do (e.g. Anderson-Hsieh et al. 1992; Munro and Derwing 1999). Prosody has always been a learning problem for both native and non-native teachers. Especially non-native teachers have been shown to produce deviant prosodic features (see Wennerstrom 1998; Pickering 2004). In teacher training, the phonetic and phonological basics of L2 intonation are often neglected, which results in an insecurity for foreign-language teachers about how to identify, teach and correct deviant prosodic forms in the classroom (cf. Walker 1999; Hirschfeld and Trouvain 2007: 177; Sonsaat 2017). Vykouková (2014: 31-32) surveyed Czech teachers of English and found that the majority (68.9%) usually does not teach pronunciation systematically and without systematic planning and previous preparation, and only 54% of the teachers reported to teach pronunciation in every lesson (see similar results by Walker 1999 with EFL teachers from Spain). Non-native teachers and lecturers at the university level are often rated as less competent due to their foreign accent, and students are, therefore, more likely to switch classes to one taught by a NS (cf. Volín et al. 2018: 79). In Czech, German, as well as Spanish schools and universities, the target pronunciation of English is predominantly Southern Standard British English (SSBrE),[4] also known as Received Pronunciation (RP) even though AmE also has an increasing influence on language models because of the prevalence of American culture in the entertainment sector (see discussion in Biersack 2002; Mompeán González 2004; Henderson et al. 2012; Jakšič and Šturm 2017). Dalton-Puffer et al. (1997: 120) explain that RP is preferred mostly because of its geographical proximity of European countries and that most students may be more likely to visit the UK. With most EFL teachers in these three European countries being NNSs, it is unclear in how far the teachers themselves approximate to the target models and they may provide a non-native accent instead as a language model for their learners. Walker (1999: 26-27), who investigated the attitudes of EFL teachers towards their own accents; found that only 2% of teachers would opt for an intelligible accent that is recognizably non-native, i.e. Spanish in this case.

English prosody has been described as “seemingly impossible to […] learn” in an L2 as an adult (Chun 1998: 74), and many teachers think that it is “not teachable, and possibly not learnable either” (Taylor 1993: 2). This is further complicated by the scarcity of exercises on prosody in English teaching materials and books in schools. For an evaluation of the inclusion of intonation in 12 English text books see Derwing et al. (2013), who conclude that intonation is mentioned many times, but the patterns presented are quite limited (e.g. only intonation rules in yes/no questions and lists are shown). Additionally, since most research articles on L2 prosody are written in a very academic style, this might complicate the understanding for teachers of English of such materials. As a consequence, “[t]he teacher fails to teach it, and the learner fails to learn it. Like other elements of language, some gifted learners will pick it up more or less unconsciously; but many will not” (Wells 2006: 2).

Additionally, while NSs acquire prosodic properties from their parents, i.e. via child-directed speech (CDS) containing exaggerated emotional prosody, they do not seem to use their prosodic patterns as much when speaking to foreigners. When NSs talk to foreigners, they mostly focus on hyper-articulation of vowels instead of their prosodic features. Adult foreigner-directed speech (FDS; Ferguson 1971) might be prosodically simplified and, thus, NNSs might be exposed to prosodic patterns even less when interacting with NSs (e.g. Biersack et al. 2005, who found FDS to exhibit a slightly narrower f0 range compared to adult NS directed speech). Other studies have found FDS to be characterized prosodically by

[4] SSBrE is the most recent term to refer to RP.

exaggerated prosody, i.e. a higher and wider f0 range (cf. Smith 2007; Romero-Trillo 2019).[5] Thus, NSs might adopt different strategies in their prosodic production when speaking to non-natives. These strategies might include entrainment towards the learners' performance, maintenance of their normal prosodic patterns (e.g. Knoll and Uther 2004; Biersack et al. 2005 (based on imaginary situations); Knoll et al. 2006; Uther et al. 2007; Knoll and Costall 2015), the dropping of most prosodic features and focus on segmental ones, or the use of more emphatic speech (exaggerated intonation, speak louder and slower). NSs, thus, adapt their speech to address the particular needs of their interlocutors, which might involve prosodic modifications to their speech. So, in a natural informal setting learners might or might not be exposed to natural native prosody, and some learners might or might not pick it up from NSs. Therefore, the foreign language classroom is one of major opportunities for all learners to practice and improve their prosodic patterns. However, foreign language teachers are rarely experts in prosody and do not usually receive training or support from teaching materials.

One of the sources teachers consult for guidance in their L2 teaching is the *Common European Framework of Reference for Languages* (CEFR)[6] (Council of Europe 2018: 136). The new CEFR descriptors mention prosodic features in the "phonological control" grid for the A1-C2 abilities:

C2	Can exploit prosodic features (e.g. stress, rhythm and intonation) appropriately and effectively in order to convey finer shades of meaning (e.g. to differentiate and emphasise).
C1	Can produce smooth, intelligible spoken discourse with only occasional lapses in control of stress, rhythm and/or intonation, which do not affect intelligibility or effectiveness. Can vary intonation and place stress correctly in order to express precisely what he/she means to say.
B2	Can employ prosodic features (e.g. stress, intonation, rhythm) to support the message he/she intends to convey, though with some influence from other languages he/she speaks.
B1	Can convey his/her message in an intelligible way in spite of a strong influence on stress, intonation and/or rhythm from other language(s) he/she speaks.
A2	Can use the prosodic features of everyday words and phrases intelligibly, in spite of a strong influence on stress, intonation and/or rhythm from other language(s) he/she speaks. Prosodic features (e.g. word stress) are adequate for familiar, everyday words and simple utterances.
A1	Can use the prosodic features of a limited repertoire of simple words and phrases intelligibly, in spite of a very strong influence on stress, rhythm, and/or intonation from other language(s) he/she speaks; his/her interlocutor needs to be collaborative.

Even though the importance of intonation is highlighted for each CEFR level, there are no specifics on how one could assess these features in the EFL classroom. Therefore,

[5] See also Papoušek and Hwang (1991) in Mandarin Chinese speech, which showed a greater f0 range in FDS, however, while lexical tonal information was heightened, suprasegmental information was reduced in FDS.
[6] Companion volume with new descriptors.

Cauvin (2013: 85) provides the following suggestions: at the B2 level f0 range in exclamatives should not be lower than 6 semitones (STs) (fail), 9 STs would be a pass, and 12 STs would be a distinction grade. Cauvin (2013) also suggests creating 'learner profiles' to analyze the learners' prosody. These suggestions will be revised in Section 5.6.3.

Despite the neglect of prosody in EFL classrooms, the present study will test whether the highly "advanced learner variety" (cf. Callies 2008) of English described here exhibit a native-like prosody.[7] Ortega and Byrnes (2008: 7) list the defining features of advancedness, i.e. institutional status, information from standardized tests, late-acquired language features, and the sophisticated use of language in context. The students described in this study rank as above average in terms of their talent for languages, their intelligence, motivation, and their attitude towards improving their English language skills, which have been shown to be crucial factors for successful SLA and learning (e.g. Lightbown and Spada 2013). As will be shown in the next Section (1.3), a great majority of the German learners opt for a native-like target norm, instead of a *lingua franca norm* (cf. Mukherjee and Rohrbach 2006). The same can be assumed for Czech and Spanish learners of English. Therefore, the students included in the present study can be expected to exhibit an almost native-like or near-native-like competence in their English and should, thus, reveal quite subtle deviances in comparison to the NS groups. The present study, therefore, investigates the productions of a very small and specialized group of L2 speakers of English, potentially rendering it not fully applicable to the average Czech, German, or Spanish L2 speakers of English. This point will be discussed further in Section 3.8.

1.3 The native-non-native speaker continuum

First of all, it is important to define the terms *native* and *non-native speech* as it will be used in the context of this study. Non-native speech is often also called *interlanguage* (IL) (Selinker 1972: 214), which according to Selinker (1972) can be described as the observable output which results from a learner's attempted production of a target language (TL)[8] norm. According to Selinker (1972), the notion of IL introduces learner language as a variety in its own right, which can be regarded as a systematic performance rather than comprising random errors. In this respect, Corder (1967) proposed that properties of L2 speech that deviate from native speech should not be considered as errors but instead be investigated as evidence for the language learners' cognitive processes. Furthermore, Selinker (1972: 221) states that many of these IL linguistic structures persistent in L2 speech are *never* really eradicated for most L2 learners, and that they rather reappear in their productions on a regular basis, especially under conditions of anxiety and with new subject matters, these linguistic structures are also called *fossilization*. The IL of EFL speakers can be distinguished from native varieties and *English as a Second Language* (ESL) *varieties* mainly in terms of the acquisition process (cf. Steinberg and Sciarini 2006: 134-135). While NSs of, for instance, BrE, AmE, or Australian English (AusE) acquire the language from early childhood, NSs of ESL varieties such as Indian English or Sri Lankan English, acquire the language in an English speaking environment along with other languages (ibid.). L2 learners of English, however, acquire the

[7] For a discussion on advanced L2 phonology see Archibald (2018).
[8] The target language (TL) is the L2 the learner is attempting to approximate (cf. Selinker 1972: 213).

foreign language in a non-English-speaking country and in the classroom (ibid.). Nevertheless, it is worth mentioning that the boundaries between ESL, EFL and *English as a Native Language* (ENL) are rather fuzzy.

As described above, a *native speaker* is defined as an individual who acquired his/her mother tongue from early childhood. Even though it might seem like a straightforward endeavor to differentiate native and non-native speakers, there is, however, huge within-speaker variability. For this reason, it is difficult to pin down the concept of the "ideal" model NS, since every NS is "stylistically multilingual" (cf. Lyons 1981: 27) and any individual NS's language use and intuitions about his/her native language will differ from NS to NS (cf. Mukherjee 2005: 13). Lightbown and Spada (2006: 202) define a NS as "[a] person who has learned a language from an early age and who has full mastery of that language". The term *native speaker* refers both to the individual speakers and to an "abstract level of language competence" (cf. Mukherjee 2005: 9). Generally, there are three central aspects of what we could describe as a native-like performance in a certain language: (1) *lexico-grammaticality*, i.e. what is possible in the language in terms of its lexis and grammar; (2*) acceptability*, i.e. appropriateness of language depending on the context; (3) *idiomaticity*, i.e. knowledge of linguistic routines (cf. Mukherjee 2002: 75, 80). A NS would have great knowledge about all three areas mentioned above, while a NNS would not have access to such knowledge, at least not to the same extent.

There have been several critiques of this native-non-native speaker continuum. Many challenge the plausibility and relevance of the concept of the NS from a linguistic and language-pedagogical point of view (cf. Paikeday 1985; Kachru 1994; Widdowson 1994: 385; Kramsch 1998: 27; Piller 2001; Birdsong 2005; Mukherjee 2005: 7). Even though I agree with Piller (2001: 112) on the point that the conception of the NS is a myth and that a "native speaker of Standard English is logically impossible", for linguistic research, it has been argued that the NS concept "indeed is a myth but a useful myth" (Davies 2003: 214). Even though prosody is highly context-dependent and there is considerable regional variation, prosody can be learned and most of the learners in the present data set have been to different English-speaking countries and might have acquired prosodic features during their stays abroad. In addition to that, learners themselves prefer either BrE or AmE as their target model of English pronunciation, which will be discussed in the following paragraph. For these reasons, I adhere to a modern definition of the NS and the use of a "native-speaker benchmark" (Foster and Tavakoli 2009: 867) in corpus-based *Contrastive Interlanguage Analysis* (CIA; cf. Granger 1996, 2015) studies. Besides these critical opinions towards adhering to NS norms in ELT, there is also the debate about the 'negativity' of the term "non-native" speaker itself. Instead of using the term "non-native speaker", different terms such as "second language user" (Cook 1999: 187-188) and even "emergent-bilingual" (García 2009: 322), have been proposed. While I acknowledge the discussion and the proposed terminology, I adhere to the terms native and non-native speech in the present study.

Furthermore, linguistic approaches to the NS norm are either *usage-based* or *usage-independent*. The present study adopts a usage-based definition of the NS norm, since it is based on the NSs' actual language use and does not draw on "idealizations and gross oversimplification of the language which native speakers are *believed* to use" (Mukherjee 2005: 9, emphasis in original). By doing so, I will describe actual NS language use by means of corpus-linguistic analyses and, therewith, it becomes possible to describe what Mukherjee

(2005) labeled an “abstracted corpus norm” (ibid.: 15). This norm can, in turn, provide us with insights into the SLA processes of advanced Czech, German, and Spanish learners of English. Therefore, I agree with Mukherjee (2005: 20), who points out that the NS concept in linguistics and language pedagogy should be retained because it shows which variety of English the learners want to approximate to. Few studies have investigated the attitudes of the learners towards varieties of English and their preferences for reference accents,[9] e.g. CzE speakers (e.g. Jakšič and Šturm 2017; Brabcová and Skarnitzl 2018, analysis with linguistic (other than English) and non-linguistic learners of English), SpE speakers (e.g. Cenoz and García Lecumberri 1999; González Ardeo 2003; Mompeán González 2004; Lasagabaster and Sierra 2005; Lasagabaster 2005; Ibarraran et al. 2008; Nowacka 2012; Carrie 2017), and GerE speakers (e.g. Dalton-Puffer et al. 1997; Mukherjee 2005; Edwards and Fuchs 2018).[10] These studies show that the NNSs in the European context recognize the need for a reference accent and find the native models of BrE (and some AmE) the most useful and necessary ones. However, the attitudes towards the two most common varieties, i.e. AmE vs. BrE differ. These studies mentioned above (at least those that specifically asked the question) show that the majority of learners from all three countries seem to prefer the RP, calling it prestigious, sophisticated, elegant, majestic, pure, standard, authentic, original, correct, and original accent are some of the words used to describe it. Frequent adjectives to describe GA are modern, widespread, socially attractive, and influential (Mompeán González 2004: 1047; Jakšič and Šturm 2017: 364). Carrie (2017) however reports that SpE learners have greater feelings of solidarity towards GA speakers. The native model does not only seem to be an ideal for students of English or other subjects, but also for language teachers themselves (cf. Walker 1999). Thus, the motivation to prefer one accent over the other seems to be socially- and culturally-motivated, rather than linguistically-motivated. However, Carrie (2017: 442), who analyzed Spanish speakers’ preferences on which accent is easier to learn, most participants agreed that a GA accent was easier for Spanish speakers to achieve, even though they prefer an RP accent. A recent unpublished survey of 111 students at the Justus Liebig University Giessen, conducted by the author of the present study, revealed that AmE is slightly more often preferred than the BrE accent. Generally, however, the BrE accent seems to be the preferred accent in most studies, but AmE is gaining more and more importance and influence due to “films, television, popular music, the Internet and the World Wide Web, air travel and control, commerce, scientific publications, economic and military assistance, and activities […] in world affairs” (Algeo 2010: 183). Brabcová and Skarnitzl (2018) analyzed 145 Czech respondents who studied other disciplines than English. The analysis showed that 70% of the CzE speakers aspire to a native-like accent (most frequently General BrE). Most of the speakers in the survey indicated to use English with foreigners or other Czechs but not with NSs of English. Interestingly, female speakers seem to prefer the BrE accent while Czech males seem to be also quite fond of Celtic accents (i.e. Irish and Scottish English) besides the BrE accent (Brabcová and Skarnitzl 2018: 45). Mompeán González (2004), found a similar

[9] To refer to the speech of an ‘ideal’ native speaker, different terms have been used, i.e. norm, target, goal, model, and reference accents (see Dziubalska-Kołaczykand Przedlacka 2008). In the present study I will not differentiate between these terms and they will be used interchangeably.

[10] This applied mostly to English language users who had a positive attitude towards English in their country. These participants stated that they do not wish for their accents to be recognizably German (Edwards and Fuchs 2018: 661).

result in that MurE speakers (survey from 2002) also prefer an Irish accent (12% of the university students), which originates from them "having visited Ireland, or their love for Ireland and Irish culture" (1047). Brabcová and Skarnitzl's (2018) results suggest that the ELF approach and the *Lingua Franca Core* (LFC) concept are not relevant to young CzE speakers. Thus, if even non-linguistics students or English language students who primarily communicate in English with foreigners want to sound more native-like (see also Edwards and Fuchs 2018: 661), then this most definitely applies to motivated and advanced language learners at university. In L2 prosody research (see Table 3 in Section 2.3 for an overview), almost all studies use a NS baseline to compare the productions of learners. Therefore, these studies support a need for a native model of pronunciation among L2 learners.

Additionally, the problem of comparing NNSs (multi-competence) with monolingual speakers (mono-competence) of a language has been criticized. These critics argue that it might be more realistic to drop the monolingual NS norm and rather adopt bilingualism (and/or early bilinguals, i.e. learners of a language), as a benchmark against which L2 productions and any L2 phenomena are compared (cf. Selinker 1972; Cook 2002; Ortega 2009; Zárate-Sández 2018: 930). For instance, Zárate-Sández (2018) found in his study that advanced learners of Spanish produced intonational patterns that were equivalent to Spanish-English bilingual heritage speakers, thus in between the values of English and Spanish monolinguals. Acknowledging this line of argumentation, it will be discussed in the results section how the less proficient learners compare to the more advanced ones, instead of only comparing them to the NSs of the present study. Nevertheless, in my opinion, a comparison to a NS target norm is the most central and crucial point of analysis.

Many SLA researchers contend that after a certain learning period, the development of a learner ends and a stabilized end-state is reached where no further approximation to the target norm is observable, also called "ultimate attainment" (cf. Gut 2009: 265). Several studies state that languages acquired in adulthood also tend to remain foreign languages to different degrees (cf. Bongaerts et al. 1997; Scovel 2000: 217; Mennen 2004; Pickering 2004; Mukherjee 2005: 11; Ward and Gallardo 2017). It is claimed that ultimate attainment is possible for some learners with individual profiles of early age of acquisition (AoA), high level-aptitude (Anufryk 2012), and motivation/autonomy (self-concept) (Moyer 2004, 2013, 2017, 2018). Ultimate attainment may be specified more broadly, i.e. the knowledge of the L2 phonology or syntax in general or more narrowly, e.g. the use of final boundary tones and intonational phrasing. Once ultimate attainment is reached, Selinker (1972: 215) assumes the existence of a fossilization mechanism which involves a permanent retention of deviant forms in the IL, which can also involve tonal features and are resistant to any kind of instruction or time that passes.[11] In a later publication, Selinker and Lamendella (1978: 187) point out that fossilization occurs before the learner reaches ultimate attainment in the TL. Therefore, fossilization can be seen as a direct result of L1 transfer. While some researchers (e.g. Gut 2009: 265) found no evidence for fossilization in L2 prosody, others claim the existence of prosodic transfer and fossilization (e.g. Orrico et al. 2016). In Orrico et al.'s (2016) study on prosody in yes-no-questions (read speech) with Italian learners of English with different

[11] Factors that influence fossilization processes include (among other factors) a lack of motivation, "lack of negative feedback, loss of attention to feedback, insensitivity to external input, maturational constraints and psychological and social variables" (as summarized in Gut 2009: 265; Markham 1997; see Long 2003 for an overview; Moyer 2004; see also Han 2009, 2013).

proficiency levels (beginners, intermediate, and advanced), a lack of improvement from Group A to C suggested the presence of prosodic transfer and drifts. The learners in their study had not spent a significant period of time in English-speaking countries and they were compared to a group of advanced speakers with English-language experience in the UK. Those students with experience abroad showed improvement in only some of the prosodic cues analyzed (ibid.). Orrico et al. (2016) conclude that those prosodic cues, which did not improve after a stay abroad, are more susceptible to fossilization. However, several studies have shown that advanced learners of English indeed show native-like prosodic performance and sometimes cannot be distinguished from NSs (e.g. Mennen 2004; Gut 2009; De Leeuw et al. 2012;[12] Mennen et al. 2014). Advanced learners have been identified as a good source of information to determine factors that constrain or support a foreign accent in L2 speech (cf. Moyer 2004: 43). Thereby, a "foreign accent" does not only refer to segmental features but also prosodic ones (cf. Van Els and De Bot 1987; Moyer 2013: 11). In order to be able to make suggestions and for the learner groups in this study to make improvements as far as their prosodic language skills are concerned, I will consider deviances from and approximations to their speech in comparison to the NS target norm they try to approximate to. Even though it is impossible to find out which variety of English the learners in the present study really aimed for, it will be inferred from their choice of country (stay abroad) and thus the variety they have been most readily exposed to. Those learners who have not been abroad in an English-speaking country will be assumed to approximate to a BrE target, since this is the most frequently taught variety in the Czech Republic, Germany, and Spain.

1.4 SLA theory, research, and L2 prosody

In the 1980s, discourse intonation first came within the purview of SLA research and the teaching of foreign languages (cf. Brazil 1997; Chun et al. 2008: 324; Celce-Murcia et al. 2010: 35). The acquisition of L2 prosody involves various aspects, such as accent placement, the phonetic realization of pitch accents, the contextually appropriate use of phonological types of accents, etc. (cf. Ortega-Llebaria and Colantoni 2014). The current section reviews the various internal, external, and extralinguistic factors that influence the acquisition of L2 prosody. In this section the focus lies on the learner, the L2 prosodic acquisition process, and how SLA and corpus-linguistic research have contributed to the understanding thereof. Thereby, one can distinguish between speaker- (Interlanguage: Selinker 1972; Basic Variety (BV): Klein and Perdue 1997), language- (PAM; SLM; MDH),[13] and ability-related SLA theories (e.g. Fundamental Difference Hypothesis: Bley-Vroman 1988, 1989, 1990). The present section focuses on speaker-dependent SLA theories (Section 1.5) and four language-dependent theories (PAM; SLM; MDH; LILt) (Section 1.4.3) in light of L2 prosody acquisition. Most of what is known about SLA phonological and phonetic L2 productions and SLA theory stems from studies that focus on the segmental level (cf. research by Flege and Best). The acquisition of L2 phonology seems to be more difficult to acquire than a native-like acquisition of vocabulary and grammar (e.g. Granena and Long 2012), even though grammatical and intonation errors or improvement tend to be correlated (e.g. Munro and

[12] Interestingly, in their study there was one participant who performed within the monolingual norms in both the L1 and L2, neither exhibiting signs of merging nor L1 attrition (de Leeuw et al. 2012: 18).
[13] *Perceptual Assimilation Model*, *Speech Learning Model*, *Markedness Differential Hypothesis.*

Derwing 1995: 89, L2 speakers who produced more grammatical errors also tended to produce more intonation errors). As far as prosody is concerned, Wennerstrom (2001: 249) claims that it “tends to be acquired unconsciously without any “monitoring” with learned rules”. The acquisition process of prosody has yet to be fully understood, but previous studies (see Section 2 for more details) have shown that L2 learners may acquire phonological properties of intonation earlier than their phonetic implementation. For instance, the nuclear stress rule is acquired later than anaphoric deaccenting (i.e. deaccenting given information), especially in SV sentences (cf. Nava and Zubizarreta 2009, Zubizarreta and Nava 2011), and simple tones are acquired before complex tones, while falls are produced before rises (e.g. Grosser 1997; Gut 2009). Furthermore, there is evidence of some development, despite different learner groups with different L1s demonstrating a different rate of acquisition. Over time, an increase in the length of IPs, an increase in the number of pitch accents they contain, and a change in the prevalence of certain pitch accents or tunes, and adopting more target-language appropriate patterns can be observed for some learner groups. In a qualitative investigation, Gut (2017: 214) found that the more prosodic features improve in non-native speech, the larger the tendency is for an improvement across speaking styles, i.e. reading, free speech, and retelling. Investigations of L2 learners from a variety of L1 backgrounds have shown that learners seem to make similar kinds of “errors”, which could point towards the existence of universal patterns in the acquisition of L2 prosody (cf. Rasier and Hiligsmann 2007: 44). Furthermore, the examination of L2 speech of different proficiency levels (advanced vs. intermediate or beginners) compared to native speech “makes a substantial contribution to theory building in SLA” (cf. Saito 2018: 283).

There are two different approaches to non-native phonology theory, i.e. those describing the characteristics of non-native speech at a specific point in time (synchronic) or those that describe non-native speech over time (diachronic). However, as the overview in Table 3 in Section 2.3 on L2 prosody research shows, most research is synchronic and longitudinal studies are extremely rare. Thus, only very little is known about how prosody is acquired and/or maintained in an L2. However, both approaches (synchronic and diachronic) are united by the endeavor to investigate the L1-L2 relationship in the prosodic acquisition process and outcome, and how this is influenced by extralinguistic factors such as experience in an L2, stay abroad, aptitude, etc. As detailed in Table 3 (Section 2.3), extralinguistic factors are not always considered and very often limited to one factor. L1 influence, however, is very frequently at the heart of L2 prosody research. While some studies hypothesize about L1 influence by inferring previous knowledge about the L1 from other research (e.g. Ramírez Verdugo 2006a), others directly investigate the L1 productions by recording the same learners in both their L1 and L2 (e.g. Mennen et al. 2014), or by comparing them to L1 productions of other L1 speakers (e.g. Volín et al. 2015). As will be shown in the following sections, L1 influence is a core component of speaker-, as well as language-dependent SLA theories. As Lee (2007; as cited in Gass 2013: 209) suggests, any L1 transfer research needs to include at least two groups of NNSs in order to determine transfer or developmental factors. Therefore, the present study includes three different NNS groups from three different language families.

1.4.1 Communication Accommodation Theory and prosodic entrainment

One important framework for the study of SLA, is that of *Communication Accommodation Theory*[14] (CAT; Giles et al. 1991) and the *theory of alignment* (Pickering and Garrod 2004, 2006), since the data in the present study consists of L1-L2 and L2-L2 conversations. According to the *theory of alignment*, which is based on syntactic and lexical alignment, not prosodic alignment, successful communication is achieved by linguistic alignment of interlocutors' speech during dialogue. Giles and Powesland (1997: 234) suggest that accommodation in speech might be used to increase comprehensibility in certain situations and/or as a means of gaining social approval from the interlocutor. Other relevant SLA theories will be mentioned in Sections 1.4.2 and 1.4.3. The theory of CAT concerns links between language, context, and identity. CAT postulates that interlocutors attune their communication to each other, and, in doing so, they generally employ three linguistic strategies, i.e. convergence, divergence, or maintenance (Giles et al. 1991). Phonetic convergence specifically describes the phenomenon in which interlocutors move closer to each other's pronunciation in an interaction, also called entrainment. Various names have been used for the accommodation processes described above, e.g. "accommodation", "alignment", "adaptation", and "convergence" (as summarized in Levitan 2014). Levitan (2014) uses *entrainment* as a general term which does not specify a direction. Nevertheless, in the present study I will use the term entrainment to mean convergence to describe (cf. also Edlund et al. 2009) an increase in similarity. The opposite act, i.e. when people sound more dissimilar, will be called disentrainment. The propensity of interlocutors to become more similar phonetically during interactions involves not only segmental but also prosodic changes (cf. Lewandowski 2012). One can speak of prosodic entrainment, when speakers match their prosodic features during conversation in face-to-face interactions (cf. Levitan 2014: 1). As previous research has shown f0 is a highly imitable feature (e.g. Pardo 2010; Babel and Bulatov 2012). One of the main functions proposed for such a behavior is to facilitate intelligibility between interlocutors. While CAT sees entrainment as a rather social process, where it might be best described as intentional and goal-oriented, Chartrand and Bargh (1999) posit that the perception-behavior link describes entrainment better, making it a rather automatic process. Therefore, if a listener is more apt to perceive a certain behavior, the listener is more likely to adopt the feature and to converge to it in an interaction. More in favor of the CAT view on entrainment, Bourhis and Giles (1977) have shown that speakers might be in full control of entrainment. In their view, entrainment is not only used to reduce social distance but also to use disentrainment to show dislike and to create distance between interlocutors. Whether the CAT or Chartrand and Bargh's (1999) perception-behavior link approach applies to the present study will not be investigated. Instead, the direction of entrainment, i.e. whether convergence takes place or not, and which groups are more likely to converge to another speaker, based on different L1-L2 groupings, will be of major interest.

When considering factors that determine entrainment, studies have shown that the direction of entrainment (i.e. divergence, convergence, maintenance) depends, among other factors, on social factors, such as the interactants' mutual likability scores (Schweitzer and Lewandowski 2013; Schweitzer et al. 2017), perceived mutual attractiveness (Babel 2012;

---

[14] Also known as *Speech Accommodation Theory* or *Accommodation Theory* (see Giles and Powesland 1997).

Michalsky and Schoormann 2017), positive attitudes towards a dialect group (Babel 2010; only segmental features) or voice (Babel et al. 2014), an individual's rejection sensitivity (Aguilar et al. 2016), personality traits (openness and agreeableness: Yu et al. 2013; openness and neuroticism: Lewandowski and Jilka 2019), and language talent (Lewandowski 2012; Lewandowski and Jilka 2019). Linguistic divergence has further been described on the basis of the CAT approach to signal social distance, and a closer social relationship is usually indicated by linguistic convergence. Giles and Powesland (1997: 234) speculate that people with a higher need for social approval will accommodate more than people who seek less social approval. Some studies (e.g. Bilious and Krauss 1988; Namy et al. 2002; Pardo 2006) show that the speaker's gender determines entrainment to different extents as well as on different dimensions. Besides social aspects, task and interactional aspects also play a role in the direction of entrainment, such as modality-specific factors (Schweitzer et al. 2017; Zellers and Schweitzer 2017) and task success (e.g. lexical and syntactic entrainment: Reitter and Moore 2007; lexical entrainment: Friedberg et al. 2012; prosodic entrainment: Thomason et al. 2013).

In L1-L1 interactions Walker and Campbell-Kibler (2015) argue that vocal entrainment takes place cross-dialectally (e.g. New Zealand and U.S. Midland English), and the larger the differences are across dialects the more likely entrainment becomes. However, entrainment is restricted by the baseline phonetic repertoire of a speaker. If the phonetic variants are too far outside the speaker's repertoire, vocal entrainment may be less likely. A similar argument is made by Kim et al. (2011), who found that NSs diverged from a NS of a dialect of the same language and barely converged to a non-native interlocutor. They also suggested that the non-native accent might be too far outside the NS's articulatory and representational structure. Thus, in L1-L2 interactions, phonetic and prosodic entrainment seems to depend on the status of the interlocutor, i.e. NS or NNS. Non-native speech is often less intelligible (e.g. Bradlow and Bent 2008; Sidaras et al. 2009: 3308) and is frequently associated with more negative attitudes (McKirnan and Hamayan 1984, with Spanish/English; Bresnahan et al. 2002, especially if it is less intelligible; Dragojevic and Giles 2016, with AmE vs. Punjabi English speakers in fluency experiment), with lower intelligibility having the potential to invoke negative attitudes. However, Kim (2012) found that NSs converge to both a NS with a different English dialect as well as a high-proficiency NNS. In line with Kim (2012), Lewandowski and Nygaard (2018) found that AmE NSs clearly converged (in f0, duration, and segments) to non-native Spanish-accented speech in a shadowing task, which varied by individual speakers and despite differences in attitudes towards the non-native accents. With regard to NNSs accommodating, Pickering (2001) found in an analysis of international teacher assistants (ITAs) and native U.S. instructors that the NS teachers employed tonal structures geared towards informational and social convergence in interaction with their student listeners. The ITAs did not make use of such strategies to increase comprehensibility or to show involvement (Kang et al. 2010: 556). In her study, Lewandowski (2012) found that the phonetically more talented German NNSs and those with a more "open" personality exhibited higher convergence in pronunciation towards a NS. The gender of the speakers, however, did not play a significant role in phonetic convergence (Lewandowski 2012).

When analyzing entrainment, the following aspects have to be borne in mind. Entrainment in general can occur locally and globally (cf. Beňuš 2014). Levitan and

Hirschberg (2011), who operationalized entrainment as similarity, convergence, and synchrony, found that entrainment was most visible on intensity but also in pitch, both on a local and global level. However, they also determined that local entrainment was stronger than global entrainment. When studying pitch in interactions, an *initializing approach* is desirable in which turn pitch is judged relative to pitch in preceding turns (cf. Zellers and Schweitzer 2017: 2336). Furthermore, entrainment effects were found to increase over time within interactions (cf. Pardo 2006, 2010; Levitan and Hirschberg 2011). This means that interlocutors need time to become attuned to the interlocutor's speech before they adopt specific features in their own speech. Levitan and Hirschberg (2011) call this "late-onset entrainment". Pardo (2006, 2010) found that speakers converge relatively early on in interactions (pre vs. task comparison), and the convergence increased by the end of an interaction and persisted even after the end of the conversation (task vs. post, pre vs. post comparisons). By contrast, Lewandowski (2012) found in her study that convergence did not seem to have a long-lasting effect, and the effect did not carry over from the dialogue to other speech styles (first person narrative or read speech).

The present study will analyze f0 range and will see whether some of the changes in f0 range in the native and non-native interactions can be ascribed to entrainment effects. By doing that, local and global entrainment will be analyzed. Furthermore, whether speakers converge in absolute or relative terms, along with the factor of time (progression in the interaction), will be considered. In addition, different measures of entrainment have been proposed, which will be briefly discussed in Section 3.7.1.3. Since only very little entrainment research has been conducted on L1-L2 and L2-L2 interactions, including conflicting results, only tentative hypotheses can be made. However, it can be hypothesized that NSs as well as NNSs will exhibit entrainment effects in their prosody. I also hypothesize that for all groups female and NNSs will entrain more because of a possibly higher need for social approval (cf. Giles and Powesland 1997: 234).

1.4.2 SLA theories and prosody: Internal factors

The present section describes speaker-dependent/internal factors and concepts that are of importance to the present study, i.e. *interlanguage* and the *Basic Variety* (BV), which will be described briefly in light of L2 prosody research. In the 1960s, with the emergence of *Error Analysis* (EA) by Corder (1967), who identified and classified errors in L2 productions in order to gain insights into learner's competence, came the recognition of learner language as a system in its own right. An *interlanguage* is reduced in its form and function and is further characterized by the following features: 1) it is systematic and rule-governed, 2) it is dynamic and it can change in the course of time, and 3) it is variable in the context of acquisition. Cross-linguistic studies have shown that learners employ cognitive processes and strategies such as transfer from the L1, fossilization, as well as over- and underuse, and avoidance. The more current term for "transfer" is "cross-linguistic influence" (cf. Jarvis and Pavlenko 2009), since languages L1, L2, L3, etc. influence each other and the influence is not only unidirectional but can go in any direction, i.e. the L2 can also influence the L1. These strategies result in a unique "grammar" or a learner system with a certain degree of systematicity, described as the *interlanguage* (IL) by Selinker (1972). The interlanguage and its associated strategies used by learners differ between linguistic levels (e.g. syntax vs.

phonology), its context, task, as well as speaking style. Previous research on non-native prosody has analyzed prosodic phenomena from a task, topic, stylistic, and context perspective (Gut 2009, 2017; Andreeva and Barry 2012: 283; Ortega-Llebaria and Colantoni 2014; Colantoni et al. 2016a, b; Ward and Gallardo 2017).

Another theoretical concept describing L2 speech is that of the BV (Klein and Perdue 1997). Klein and Perdue (1997) propose that languages possibly consist of unnecessary complexity and redundancy. Opposed to this, there is "a well-structured, efficient and simple form of language - the Basic Variety" (Klein and Perdue 1997: 301). In their longitudinal study of 40 learners of English, Klein and Perdue (1997) proposed four key assumptions about the BV. The first assumption is that during the language acquisition process, learners transition systematically from one learner basic variety to another, i.e. they pass through a series of systematic learner varieties. Second, learners exhibit a limited set of organizational principles, which can evolve and reorganize during time. Third, as the interlanguage concept, the BV assumes that the learner language is a system in its own right with a particular inventory of language units. Fourth, the elementary stage until the final stage should be analyzed equally. Thus, the BV theory suggests that learners with a lower proficiency level should demonstrate a lower degree of prosodic variation, conforming more to a basic variety and learners with a higher proficiency should exhibit a greater prosodic variability approaching the performance of NSs. Even though some parallels can be observed between the BV and IL theory, fundamental differences can be observed between the two theories in terms of language processing. According to Selinker (1972), the development of an IL in SLA is different from the development of first language acquisition by children, including different psychological processes: i.e. language transfer, transfer of training, strategies of second-language learning, strategies of second-language communication, and overgeneralization of the target-language linguistic material. Transfer can be defined as the influence of the L1 when using an L2, or vice versa. There can be positive (when a certain L1 unit is transferred with an identical counterpart in the L2) or negative transfer (when corresponding units in the L1 and L2 are not identical and a transfer takes place). Interference, which is a related concept, however, means that there is an influence of the L1 on the L2 (or vice versa), which leads to an error. Transfer is a central notion of the theory of Interlanguage and it contradicts the theory of the BV. According to Klein and Perdue (1997), learners develop a variety that is only simpler in structure in comparison to a fully-fledged language, which is basically error-free. How the concept of transfer relates to the case of prosody will be discussed in Section 1.5.

Even though the present study chooses the CIA approach, I do not adhere to the use of the key terminology, since prosody is highly dependent on the context, it will be quite difficult to claim a certain feature to be over- or underused. Therefore, I will speak of deviances instead. Nevertheless, there is the possibility that some systematic competence "errors" might be revealed in the analysis, although I expect to find more unsystematic performance errors (cf. Corder 1967: 166-167). Furthermore, one has to distinguish between errors that impede communication and those that do not. In the case of prosody, most deviances will be mainly classified as foreign-sounding and not erroneous. Yet, in some cases, miscommunication might occur when a falling intonation is used for a question which is not marked grammatically and no answer is given, because it has been interpreted as a statement. While these type of "errors" might interrupt the communication flow, these types of errors are usually identified quite quickly, either by the speaker or hearer. Thus, the aspects that could be measured in terms of prosody are intelligibility, degree of foreign-soundingness, and adequate situational use. The present study focuses on those prosodic features contributing to foreign-soundingness. However, once deviations have been identified and described, I will make use of interlingual and intralingual factors to explain whether these deviances stem from L1 influence or could be considered developmental errors (see Error Analysis: Corder 1967; Ellis 2008). This binary classification might prove to be difficult at times, especially when L2 prosody is concerned, because intonational acquisition is far from being a well understood phenomenon.

1.4.3 L2 speech models and prosody: External factors

The present section is concerned with external factors, i.e. language-dependent SLA theories. One of the earliest SLA theories to account for L1 influence was the *Contrastive Analysis Hypothesis* (CAH) (e.g. Lado 1957), according to which the learner forms habits from the L1 which can then be transferred to the L2. The behaviorist descent of the CAH and its basis on habit formation becomes apparent and has been frequently criticized. However, since its inception, *Contrastive Analysis* (CA) has provided important implications for following speech learning models. In fact, CA has been incorporated into the most recent speech learning models, which will be described below. CA was also responsible for introducing a major continuing theme into SLA theory, i.e. the influence of the L1 on the L2. According to the CAH, there could be either positive or negative transfer, depending on the appropriateness of the transfer. For instance, a positive transfer would occur if learners transfer their yes/no-question intonation with a rising tone at the end, which is the same in the L1 as in the L2. An example of negative transfer would be if intonation patterns differ in the two languages and the learner transfers the L1 pattern to the L2. It is assumed in CAH that systematic comparisons of the L1 and L2 would allow for predictions of L2 errors. Thus, according to Lado (1957), L2 structures that exist in the L1 with the same form, meaning, and distribution will be the easiest to acquire. L2 structures which are different and non-existent in that form in the L1 need to be learned and the learner will have more difficulty in acquiring these forms (cf. Lado 1957: 59). Lado further stresses that the L2 structures that are the most difficult to learn are those with a partial overlap but do not have equivalence in form, meaning, and/or distribution (cf. Saville-Troike and Barto 2017: 37-38). Even though the CAH did not originally discuss prosodic similarities and differences, previous research suggests that prosodic phenomena, just like segmental ones, may also transfer from L1 to L2 and that prosody may be even more susceptible to transfer because of its abstract nature and generality (Vogel 1991: 55). L1 influence on prosody is widely attested. For instance, there is growing evidence that the f0 range might be influenced by a speaker's L1 (e.g. Backman 1979; van Bezooijen 1995, with inter-gender differences Dutch/Japanese; Scharff-Rethfeldt et al. 2008; Gut 2009: 244-245, the f0 range in reading passage style of German, Spanish, and Russian learners of English). When it comes to prosodic transfer specifically, one has to consider that transfer can be discovered both at the phonological (resulting from different tone inventories and meanings assigned to them) as well as phonetic level (e.g. realization of a phonologically identical tune) (cf. Mennen 2007: 57). Graham and Post (2018: 10) also lend support to the CAH by claiming that the typological similarity of the source and TLs (i.e. Spanish and

English) may lead to a faster acquisition in late bilinguals compared to typologically more different languages (i.e. English and Japanese).

However, previous empirical research has found that it is extremely difficult to predict errors in an L2 by simply comparing the L1 and L2, since errors are often reported that could not possibly arise from L1-L2 differences. For instance, evidence against L1 influence on prosody is outlined in various studies (e.g. Grosser 1997; partial evidence against and for: Mennen et al. 2014; Volín et al. 2015). Volín et al. (2015) finds the f0 range in Czech English to be even narrower than that in native Czech. They conclude that "implicit uncertainty" or "moderate anxiety" when speaking in the foreign language might be responsible for such effects in the f0 range (cf. Volín et al. 2015: 121). While Mennen et al.'s (2014) German learners of English show intermediate values in the f0 range between the L1 and the L2, which they interpret as L1 influence, on the whole the learners are rather successful at attaining a native-like f0 range. Mennen et al. (2014: 324) explain their results by stating that the learners in their sample might be so successful because of their level of proficiency, since they live in the UK. Additionally, task effects have been shown to influence L1 transfer in Ortega-Llebaria and Colantoni (2014), who find more L1 transfer in tasks that require a higher level of prosody processing (i.e. form-meaning mapping) than tasks with lower levels. Other studies demonstrate that cross-language similarities do not always lead to language transfer but rather to avoidance of particular structures that most closely resemble their L1 (e.g. Hulstijn and Marchena 1989, with Dutch learners of English with phrasal verbs). Avoidance, however, is hard to analyze in L2 prosody, just like segments, suprasegmentals are quite hard to avoid. On the other hand, learners may avoid prosodic variation altogether and, thus, come across as more monotonous due to their lack of prosodic knowledge/experience in the L2 or insecurities in sounding too foreign. In sum, previous research that repeatedly found counter arguments for the initial CAH, led to the more currently used term of "cross-language influence", which in its changed form can still use the premises of the CAH, which can help to explain the occurrence of L2 errors.

Since the emergence of the CAH, SLA research and theory-building have come a long way, and current L2 phonological acquisition models focus on the constraints the L1 places on L2 phonology, e.g. the *Perceptual Assimilation Model* (PAM; Best 1995), *Perceptual Assimilation Model* for NNSs (PAM-L2; Best and Tyler 2007), *Speech Learning Model* (SLM; Flege 1995), and *Markedness Differential Hypothesis* (MDH; Eckman 1996). These SLA models primarily focus on L1 influence as it is seen perceptually by the L2 learner. In addition, Eckman's MDH, which was an attempt to compensate for the weaknesses of the CAH, predicts that not all linguistic areas of L1-L2 differences will pose the same degree of difficulty, and more marked features would be acquired later and are more difficult to acquire than unmarked versions (cf. Eckman 1977: 329). By accounting not only for interference errors but also for developmental errors, the MDH contributed hugely to the understanding of interlanguage phonology (see Rasier and Hiligsmann 2007 for experimental support for the MDH and prosodic transfer). Based on phonological theory of markedness, Eckmann (1977) proposed with his MDH that unmarked phonological features are more common than marked ones in a given language. Despite a variety of SLA theories, there is no comprehensive model of non-native speech that includes all aspects concerning the acquisition of the form and function of prosody. In fact, most of the models proposed in SLA, e.g. SLM, PAM, PAM-L2, and the MDH (not restricted to the phonological level), are based on segmental rather than suprasegmental features of L2 speech, and rely on L1/L2 interference to predict segmental errors. Since all these major speech learning models are based on segmental, rather than suprasegmental features, they will not be discussed here in detail (see Gut 2009 for a review of current speech learning models). An overview of the major assumptions of the influential theories in SLA, i.e. SLM, PAM-L2, and LILt is provided in Table 1. The reason why these SLA speech models are, nevertheless, important in the context of the present study is that there is an interaction between the segmental and prosodic level and this needs to be incorporated and tested with the help of these L2 speech models. Previous research by Trofimovich and Baker (2006), for instance, suggests that L2 learning on the segmental and prosodic level share many similarities since they both improve gradually as a result of experience and exposure to the L2. Additionally, it has to be noted that all speech learning models have a particular "listener" in mind, i.e. the SLM has focused on *experienced* listeners, while the PAM has focused on primarily naïve non-native listeners. For this reason Best and Tyler (2007) proposed the PAM-L2, which focuses on two non-native listener groups, i.e. inexperienced vs. L2 speech learners.

Partially due to the absence of an established SLA model for L2 prosody, very little is known about how prosody develops in L2 learners. A first approach to standardize variables affecting the acquisition of prosody was introduced by Mennen (2015), who proposed the *L2 Intonation Learning Theory* (LILt) "working model", which means that it is subject to change as more data is published on non-native intonation. Mennen's (2015) LILt model is the only speech learning model so far that exclusively accounts for L2 prosodic perception and production difficulties and formulates predictions where prosodic deviances are likely to occur. For a direct comparison of other L2 speech learning models for segmental features and how it can be integrated in the LILt model, the reader is referred to Mennen (2015). Therefore, more research is needed to test Mennen's (2015) LILt working model. The present section will, thus, present Mennen's (2015) LILt model and show how it compares to general assumptions of L2 models of segmental speech learning (i.e. SLM, PAM, and MDH). Mennen's (2015) LILt model includes four dimensions (modified from Ladd 1996) along which languages are thought to differ intonationally:

1. the inventory of structural *phonological* elements (e.g. pitch accents and edge tones) ('systemic dimension')
2. the *phonetic* implementation of these structural elements (e.g. alignment of pitch accents with segments) ('realizational dimension')
3. the distribution of boundary tones and pitch accents (e.g. frequency of use and combinations of pitch sequences, intonational phrasing) ('distributional dimension')
4. the functionality of these structural elements (e.g. how they are used to signal linguistic functions such as focus or interrogativity) ('functional dimension')

Dimensions one to three are understood as corresponding to the structural aspects of prosody in a language (cf. Mennen et al. 2010: 320). Thus, if an L2 learner is determined to master the prosody of a language, he/she needs to learn the structural elements, like pitch accents, how these are realized and distributed in the respective language, and how they are used to fulfill different communicative functions in different contexts. The present thesis includes all four dimensions when examining L2 intonation deviation. Mennen (2015), who reviewed L2 intonation studies in respect to the four dimensions proposed in her LILt model, shows that deviations in L2 speech have been found on each of the four dimensions; however, some dimensions appeared to be more susceptible to deviation than others. Her review demonstrates that the realizational dimension of intonation might be the one that exhibits the most deviations from the target norm; while the systemic dimension does not seem to cause many problems, the semantic dimension has also proven to be difficult for learners to produce (see Mennen 2015 for an overview of these studies). Furthermore, Ward and Gallardo (2017: 22) find that the Spanish NNSs of English also indicate a weakness with their prosodic expressions and their pragmatic functions associated with them.

Gut (2009: 300) analyzes the speech of 101 NNSs (including German, Spanish, Russian, and other unspecified L2 speakers of English) with regard to its segmental, suprasegmental, fluency, morphosyntactic and lexical features in relation to various non-linguistic factors in three different speaking styles and develops a multifactorial model depicting the interlanguage of the learners as consisting of separate but partly connected domains of phonology, morphosyntax, lexicon, accent, and fluency. In her model, intonation is strongly influenced by fluency and vice versa. Her analysis illustrates that proficiency in one of these domains cannot predict proficiency in another. Gut (2009: 302) finds these speaking style and task-dependent differences in intonation and other phenomena to be greater than the differences resulting from different L1 backgrounds, learning contexts, and proficiency levels. Additionally, there is no evidence for universal prosodic and segmental features in the non-native speech samples (cf. Gut 2009: 302). However, the learners come from different L1 backgrounds and are not systematically separated in her analysis. She concludes that a combination of the non-linguistic factors age of learning (AoL), length of residence (LoR), instruction, knowledge, and motivation contribute to the shape of L2 phonology (cf. Gut 2009: 302).

Table 1. Comparison of selected language-dependent SLA theories

Predictions	SLM (Flege 1995)	PAM-L2 (Best and Tyler 2007)	LILt (Mennen 2015)
Perception - Production	• Concerned with non-native production & perception • Claims that many difficulties may be perceptually motivated, i.e. inaccurate perception of sounds in an L2 may cause inaccurate production	• Concerned with NN speech perception • Claims that many difficulties may be perceptually motivated • Equivalence between L1 & L2 at the lexical-functional level may play a role	• Concerned with NN production & perception • Many difficulties may be perceptually motivated • Necessary to consider both form (realizational dimension) & meaning (semantic dimension) when predicting relative difficulty of L2 intonation categories
L1/L2 Similarity/ Difference	• Perception of L2 segments depends on similarity of phonetic properties of L2 segment & L1 categories • When L2 segments are sufficiently similar to L1 categories, they will be 'equivalence-classified' to L1 categories & deviances in production are likely to occur • When L2 segments are sufficiently different from L1 categories, the development of a new category should be possible	• Perception of L2 segments depends on the similarity of phonetic properties of the L2 segment & L1 categories • When L2 segments are sufficiently similar to L1 categories, they will be 'perceptually assimilated' to L1 categories and deviances in production are likely to occur • When L2 segments are sufficiently different from L1 categories, the development of a new category should be possible	• Recognizes that similarities and dissimilarities between L1 and L2 intonation can occur along more than just the systemic dimension • Learners' perception of intonational cues that are not present in or differ from the L1 is often poor • Variation in the realizational dimension may impact learner's ability to discriminate, categorize, & produce L2 phonological categories • The position and context in which certain contrasts occur is equally important in intonation, and needs to be tested & controlled for
Age of Acquisition /Arrival/ Exposure	• Age of arrival (AoAr) or AoL as important predictors of success • SLM-r[15] (Flege 2018) L2 learners of all ages can learn to produce and perceive L2 phonetic segments accurately (if they have enough input)	• AoAr or AoL as important predictors of success • Same basic perceptual learning abilities are available to adults learning an L2 as to children learning an Ll or L2	• AoAr or AoL in L2-speaking country is an important factor in predicting overall success in acquiring L2 intonation • "Earlier is better" also applies to L2 intonation learning, but influence of AoAr is not necessarily the same for each dimension of intonation
L2 Experience Proficiency	• Substantial amount of input required for development of accurate information about phonetic properties of L2 speech sounds & establishment of phonetic categories • With increasing experience/exposure, learners become more "perceptually attuned" to L2 sounds & may become more native-like in production	• SLM and PAM-L2 agree that it should be possible for learners to ultimately approximate or reach L2 norms in production with increasing experience/exposure to the L2	• Increasing experience in L2 will lead to more native-like production of L2 intonation • Learners with limited experience will rely on their L1 structures and L1 transfer will take place • Not all intonation dimensions constitute same amount of difficulty in L2 learning. Ultimate attainment may be possible in some intonational dimensions
L1 - L2 Bi-directional Interaction	• L1 and L2 categories exist in a "common phonological space", which may cause languages to interact (can be bidirectional) • Interaction between L1 & L2 can take the form of assimilation or merging of L1 & L2 properties, where L2 learners tend to produce intermediate values • Dissimilation used to maintain phonetic contrast between categories in L1 & L2	• PAM-L2 agrees with SLM that L1 and L2 phonological categories exist in a common space • PAM diverges from SLM in positing that both phonetic *and* phonological levels interact in L2 speech learning, and that this depends crucially on the relationship between the phonological spaces of the L1 and L2	• The notion that L1 and L2 categories exist in a common phonological space and an interaction between them is compatible with LILt's viewpoint • Interaction effects have been found for intonation, i.e. merging effects & overshooting • Further research is needed to clarify what factors govern assimilation and dissimilation effects, and why some speakers are able to entirely maintain or achieve separateness of L1 & L2 systems

[15] The *r* stands for revised version of the SLM (see Flege 2018).

Thus, in order to predict and explain the areas of difficulty in L2 prosody, the learners' age (younger vs. older learners) will be analyzed in the present study, as a factor of native-like performance in the L2 concerning the four intonational dimensions. Among other factors analyzed and discussed in Sections 3 and 4, one can partially infer the learners' level of proficiency[16] with the age, the number of months in an English-speaking country, and we can deduce the experience/exposure to an English native variety, and thus test whether any of the prosodic features within the four intonational dimensions reach target-like values. However, the interaction effects (assimilation and dissimilation) cannot be analyzed in the present study, since L1 recordings of the L2 speakers do not exist.[17] By doing this, the present study contributes to the assumptions of Mennen's (2015) working model, as well as the other SLA theories outlined in Table 1. The objective of this study is to contribute to the development of a unified model of L2 prosodic acquisition by investigating the productive prosody of learners of typologically different first languages. In particular, the present thesis adds empirical results to the questions posed by Mennen (2015: 184) within the LILt framework:

1. Are deviations equally reflected in different dimensions of intonation?
2. Are some intonation parameters more susceptible to transfer than others?
3. Do deviances in different dimensions of intonation diminish in parallel?
4. Are there symmetries in the pace and trajectory across learners of different L1 backgrounds?

The present study aims at proving or disproving some of the aspects of current L2 acquisitional theories (more specifically the LILt framework), and attempts to contribute empirically to a more comprehensive theory of intonational L2 acquisition. The aim is primarily to describe L2 prosodic features and to evaluate various explanatory factors for the present data set. Mennen's (2015) predictions will be tested in the present study, as far as the available metadata on the learners allows.

1.5 Extralinguistic factors influencing SLA

Empirical research on L2 prosody within SLA theory has shown that not all deviances from the native target norm can be explained by L1 influence. Deviances for which no L1 transfer can be attested are usually termed developmental, and this is when speaker-dependent extralinguistic factors are considered in finding explanations. These extralinguistic factors that can influence learner performance are age (AoA and testing), experience (quality and quantity of input), length of residence (LoR), proficiency (usually measured by a test or ratings), and some that are quite difficult to measure such as motivation, personality, and aptitude. It is likely that these factors (among others) interact with each other and influence L2 production; however, in most cases these interpretations remain speculative (cf. Jilka 2007: 85).

[16] Criticism of such an approach is discussed in Section 1.5.3.
[17] Additional 31 L1 recordings of the LINDSEI-CZ corpus were collected but are not available to the general public (see Gráf 2019).

1.5.1 Age

There is disagreement on the role of AoA when it comes to L2 prosody (e.g. Trofimovich and Baker 2006; Mennen et al. 2014). It is assumed that after a so-called *critical period*, it is much harder for language learners to fully acquire an L2, due to the loss of cerebral plasticity, especially its phonology, including prosody. The critical period has been set at different age cut-off points by different researchers (10-12: Lenneberg 1967; Seliger 1978; six to eight: Walsh and Diller 1981; Scovel 1988). These multiple critical periods are mainly linked to the acquisition of a certain level of linguistics, and the acquisition of syntax and morphology declines after the age of 15 (e.g. Patkowski 1980; Long 1990) and the acquisition of phonology declines much earlier at six to twelve years of age. Recent studies (Hartshorne et al. 2018; Trafton 2018) also suggest that the grammar of a language can be easily acquired up to the age of 17 or 18; however, they also agree that a native-like proficiency is nearly impossible if a language is learned after the age of ten. Previous research suggests that the ability to correctly replicate the L2 patterns declines earlier for intonation than for pronunciation and if an L2 is acquired between even to eleven years, a slightly foreign intonation will be perceived by raters, and after 12-13 years a marked L2 accent will be perceived (e.g. Tahta et al. 1981a, b). After the late 20s cognitive functions, such as working memory, attention, and speech sound processing, start to decrease (Birdsong 2014 as cited in Saito 2018: 294). Therefore, Saito (2018: 294) describes the language learning process of adult L2 learners as “rather explicit, effortful, and incomplete”. In contrast to these critical or sensitive period stances, Hyltenstam and Abrahamsson (2000: 157, 2003) propose a *Linear Decline Hypothesis* (LDH) in second language performance (reflected in the view of Birdsong 1999: 12), which persists throughout a person’s lifespan (contrary to the *Critical Period Hypothesis* (CPH), which implies a non-linear relationship between age and L2 ultimate attainment).[18] Thus, age effects do not occur at or after specific periods but rather just reflect a general aging process. Related hypotheses in line with the main postulate of the LDH, are the “exercise hypothesis” and “single system hypothesis”. The “exercise hypothesis” (Bever 1981; Johnson and Newport 1989: 64; Hurford 1991; Flege et al. 1997b: 171) places an emphasis on continually learning a language uninterruptedly and, thus, the ability to produce and perceive L2 speech remains intact across the life span. Similarly, Flege (2019) argues for a non-critical period, which is defined by age-related variation in L2 input, where the quantity and quality of input plays a very important role in ultimate L2 attainment. These hypotheses are, however, extremely hard to test, since such detailed longitudinal information is not available on learners in existing L2 speech corpora. The “single system” hypothesis (cf. Flege et al. 1997b: 171-172) postulates that discontinued L1 use may reduce the extent of a perceived foreign accent when speaking/learning an L2 (for empirical studies see Piske et al. 2001 and Flege et al. 1997b).

Studies on how age influences the development and production of prosodic features are quite scarce. Therefore, the following results have to be interpreted with caution. Some support for the LDH was found by De Leeuw (2019) in a longitudinal study, which lasted for four decades, of the speech of Stefanie Graf, a late German-English bilingual. De Leeuw

[18] For a further discussion of neural plasticity of phonetic and phonological features see de Leeuw and Celata (2019).

(2019) found that her German prosody significantly decreased in f0 level and span as an effect of increasing age, but so did her average f0 maximum increase over time, which De Leeuw interprets as a result of English L2 acquisition of prosody. Trofimovich and Baker (2006), who analyzed the production of prosodic (stress timing and peak alignment) and fluency features (speech rate, pause frequency, and pause duration) in declarative utterances by adult Korean learners of English, showed that the AoAr in the U.S. (and thus their first extensive real-life exposure to the L2) influenced only the production of the three fluency features and had no effect on the prosodic features analyzed. Interestingly, even an earlier start of language learning does not always guarantee accent-free speech in adulthood (e.g. Flynn and Manuel 1991; Flege et al. 1997b; Flege et al. 2002; Flege and MacKay 2004; Uzal et al. 2015). Nevertheless, an array of studies cross-linguistically report on exceptional learners who have reached “ultimate attainment” and are hardly distinguishable from NSs when it comes to their pronunciation, despite a late start in L2 acquisition (e.g. Bongaerts 1999; Bongaerts et al. 2000; Gut 2009; Ward and Gallardo 2015, 2017). The outstanding performance of these learners have been attributed to a special talent in second language acquisition (Ioup et al. 1994; Bylund et al. 2012), and others have been attributed to a high motivation (Moyer 2004). Regardless of the age of the learner, various factors have been claimed to majorly contribute to successful phonological and phonetic ultimate attainment: e.g. the (strong) willingness, desire, or motivation to sound native-like (for different purposes), or ‘affinity’ for the TL and culture, and the continued effort to practice the L2 pronunciation (e.g. Suter 1976; Purcell and Suter 1980; Bongaerts et al. 1995, 1997, 2000; Elliott 1995; Moyer 1999, 2004, 2007, 2018). In Elliott’s (1995) study, the attitude towards or concern about pronouncing the TL (Spanish) was the most significant factor in the accurate TL pronunciation.

In sum, as previous research on L2 prosody and the age variable have shown, the window of learning L2 prosody does not close when an L2 is learned in adulthood, and the learner’s L1 might have a bigger influence on successful L2 acquisition than maturational aspects (cf. Lengeris 2012). While the CPH, LDH, exercise and single system hypotheses are not testable with the data set in the present study, only the age of the learners and the length of English instruction in school and university can be considered in the analysis of the production of prosodic features.

1.5.2 Sex and gender

Among the individual differences between language learners, biological sex and/or socially-constructed gender seem to have an influence on their success in language learning. Even though differences among male and female learners have been proposed in terms of language learning in general, as well as prosody (e.g. Boyle 1987; Oxford and Ehrman 1993; Ullman et al. 2007; Niebuhr 2015; Moyer 2016), not all L2 prosody studies include this variable. According to Moyer (2016: 17), “gender has so rarely been addressed in studies of L2 phonology, it is typically noted as a somewhat unexpected finding, or even an afterthought, with no clear explanation offered”. Nowadays, stereotypical views on who learns a language is still endorsed in society, and women are more inclined to learn languages because they tend to hold the belief (instilled from both the home and society at large) that they need language skills to advance in their future careers (cf. López Rúa 2006: 105-106). Women’s communicative skills are upheld because of their expected roles as mothers and “emotional supporters" in the family (ibid.). Although not specifically related to prosody, general differences that have been mentioned in previous ultimate attainment and successful language learning studies are the following: 1) female learners use more diversified language-learning strategies and they use them more often, 2) they are more motivated (especially integrative motivation as opposed to instrumental motivation), 3) they have more positive attitudes towards the country, speakers, and culture of the L2, 4) they are more interested and concerned about pronunciation accuracy, and 5) they have more fun in language learning in general (cf. Oxford et al. 1988; Ehrman and Oxford 1989; Spolsky 1989: 209; Thompson 1991: 197; Bacon and Finnemann 1992; Ellis 1994: 202-204; Dörnyei and Clément 2001: 413; López Rúa 2006; Bielska 2009; Jiang at al 2009: 489; Dewaele et al. 2016; Moyer 2016). This is reflected in the significantly higher number of women in foreign language classes and universities and the persistent belief that languages are important for future jobs among females (López Rúa 2006: 100). According to Spolsky (1989: 151), integrative motivation can be observed to result in more native-like pronunciation (see also Polat 2011). Overall, several researchers agree that female learners make greater gains in L2 phonology, which might stem from their heightened sensitivity to input, more frequent L2 use, and/or superior listening comprehension and processing skills (Bacon 1992: only perceived comprehension; Spezzini 2004; López Rúa 2006: 108; Chipere 2013; Moyer 2016: 22-23), which may be coupled with other favorable variables such as LoR and interaction with NSs (cf. Moyer 2004, 2011), age (Asher and García 1969; Flege et al. 1995: 3130), and motivation/autonomy (e.g. Moyer 2004, 2013, 2017). Summarizing L2 phonological acquisition research, Moyer (2016: 22) states that "females demonstrate a fairly consistent advantage in the phonological realm, and this is most evident in discourse-level tasks that involve suprasegmental features including rhythm, intonation, and stress". Higher motivation and language learning aptitude observed in female learners may result in more self-confidence, positive self-evaluation, lower anxiety and inhibition (e.g. Matsuda and Gobel 2004). However, female learners were also found to have more positive and negative emotions in the classroom, i.e. they worry more about making mistakes and were less confident in using the FL (cf. see also Marzec-Stawiarska 2014 who found no difference between men and women but a slight tendency for women to be more anxious; Dewaele et al. 2016). This heightened emotionality, both in terms of negative and positive emotions, however, is also correlated with successful SLA. From a biological perspective, women are more emotional, which triggers the production of more serotonin stimulated by estrogen. This has an influence on how (emotional) information is processed (Schütze 2016: 58-59). Results obtained for verbal abilities and lexical learning suggest that female learners are more successful in these areas (Kimura 1999: 101; Halpern 2012) and others claim that male learners are more successful in listening vocabulary, i.e. the comprehension of heard vocabulary (Boyle 1987), or no significant differences were found (Macaulay 1978; Hyde and Linn 1988). Prosody also has a primarily emotional function; therefore, one might predict that female learners could be more apt at acquiring more subtle prosodic differences due to their heightened emotionality and their perception of these features. In L2 foreign accent studies, gender has not been identified as a significant predictor (e.g. Suter 1976; Snow and Hoefnagel-Höhle 1977; Purcell and Suter 1980; Flege and Fletcher 1992; Elliott 1995; Piske et al. 2001), although a few studies indicate that women generally receive higher oral speaking proficiency-ratings than male participants (e.g. Thompson 1991: 189). Some studies suggest a superiority of women to speak accent-free in an L2, a small but difference (2.2% of variance) which diminishes with years of residence in an L2 country (Asher and Garcia 1969; Tahta et al. 1981a). Additionally, women have been found to be faster in recognizing familiar melodies (with and without lyrics) than men (cf. Miles et al. 2016), which might lead to the tentative assumption that women could have 'a better ear' for prosodic patterns in the L1 and L2. Miles et al. (2016) explain this result by the female advantage of storing melodies (partially) in the declarative memory; while men are thought to have an advantage when it comes to procedural memory (see Ullman et al. 2007).

In sum, it is not the socially-constructed gender or biological sex of a learner that determines successful SLA, but rather the interplay of various associated or interrelated factors with a certain sex. However, female learners seem to have certain traits that enhance their success in language learning. Neurological, cognitive (verbal intelligence and aptitude), affective (attitude, motivation, and personality), social and educational factors seem to interact with each other and they seem to be activated differently in female and male learners (cf. López Rúa 2006: 99). Personality itself can also explain the degree of foreign accentedness, and extraversion and neuroticism were found to be significant predictors of foreign accent in Zárate-Sández (2017). However, it has to be kept in mind that not all male and female speakers fit neatly into these categories. Recent research in sociophonetics has shifted away from treating subjects as either male or female to seeing gender as something that is constructed out of stances, social actions, and stylistic variation in individual contexts (cf. Cameron 2011; Eckert and Podesva 2011; Podesva 2011). While it is not possible to determine if any of the learners in the present data set have a socially-constructed gender that is not reflected by their biological sex, it is observable that some speakers' voices sound more stereotypically feminine or masculine. Another problem is that studies (the present one included) study the product (synchronic production data) instead of the process of language learning itself (longitudinal studies) (cf. Bowden et al. 2005: 112), which would be a valuable resource in finding differences and similarities between male and female speech. An example of a micro-analysis of utterance-final pitch is Podesva's (2011) analysis of three gay professionals who use final intonation patterns (falling, level, rising) with different frequencies across three different professional and social contexts. Zooming in on individual uses of f0 can inform macro categories of how gender and sexual orientation are constructed. Therefore, the use of prosodic features is to be exploited in as many individual contexts as possible. More prosody-related results on sex will be discussed in Section 2.

1.5.3 Proficiency

Proficiency level has become one of the key explanatory factors when it comes to the performance of language learners in SLA and learner corpus research. L2 proficiency describes a learner's inventory of specific linguistic skills, such as phonological, lexical, morphosyntactic, and pragmatic skills. Therefore, proficiency can be viewed as what a learner has achieved at a given point in time (i.e. time of measuring) as a result of an interplay of various factors, such as aptitude (talent for pronunciation and accents), external factors, motivation, AoL, neurological, cognitive, as well as psychological factors that are extremely difficult to control for. However, the attained level of proficiency of language learning should mirror a certain degree of language aptitude and motivation. In most cases, the proficiency of language learners are ascertained by a specific test; the levels of beginner, intermediate, and advanced (e.g. A1-C2) of the CEFR (Council of Europe 2001) are most commonly used or other tests which measure a students' receptive (listening and reading) and productive skills (speaking and writing) in English, such as the *Testing of English as a Foreign Language* (TOEFL) test. Gut (2009) has found that learners can perform well in one area of linguistics, while performing poorly in another. Lleó and Vogel (2004: 92) observed no consistent advantage for reading tasks over a spontaneous speech task in segmental learning and intonational phrasing of Spanish learners of German (even though reading tends to yield more target-like[19] results due to the level of formality involved). Only the advanced learner group showed a consistent, yet only slight advantage in the reading task compared to the spontaneous task (ibid.). Interestingly, Colantoni et al. (2016a) found that spontaneous narratives yielded more comparable and native-like prosodic patterns than read speech of Spanish heritage speakers[20] and long-term immigrants in the U.S. Since none of the existing proficiency tests specifically measure prosodic competence and only some of the learners analyzed in the present study have been tested on their proficiency levels after being recorded (cf. Huang et al. 2018),[21] I will take their age, current semester in their course of studies, and length of stays abroad in English-speaking countries as indicators for their proficiency. Section 3.3 describes in more detail the proposed proficiency levels of the learners in the present study.

As in all the factors explaining deviant prosodic features in an L2, studies come to varying conclusions on the role proficiency plays. While some studies have identified improvements (i.e. an approximation towards native-like L2 patterns) with increasing L2 input and an increasing proficiency level (Backman 1979, with Spanish learners of English; Jun and Oh 2000, with American learners of Seoul Korean; Zárate-Sández 2018: 930, with English learners of Spanish), other studies have found no effect of proficiency level (e.g. Lepetit 1989: 404, with English and Japanese learners of French). Furthermore, some studies suggest that a higher proficiency might even lead to a worse performance in the perception of intonational meaning (e.g. Chen 2009, with Dutch learners of English and English learners of Dutch). Nevertheless, the evidence for a positive effect of increasing proficiency on the improvement of prosodic features outweighs, both in L2 productions (e.g. Backman 1979; Anufryk 2012; Zárate-Sández 2018) as well as L2 accent as perceived by NSs (e.g. Flege et al. 1997b; see Piske et al. 2001 and Piske 2012 for an overview).

The overall proficiency cannot be determined for all the learners in the present data set (only CEFR levels for LINDSEI-CZ[22] and some LINDSEI-GE are available), due to the unavailability of proficiency ratings or tests. However, lexical developmental measures will also help to determine the proficiency of the learners. Bundgaard-Nielsen et al. (2012, 2011) found evidence for their proposed vocabulary-growth hypothesis, which predicts that learners with larger L2 vocabularies produce more intelligible L2 vowel sounds than those with

[19] Although based on segmental features, Piske et al. (2001) summarize that some studies (Oyama 1976: 268; and Thompson 1991) found speakers were judged to produce a stronger foreign-accent in read speech than in spontaneous speech.

[20] The heritage speakers in Colantoni et al. (2016) were educated in English and did not receive a formal education in Spanish.

[21] I want to thank Lan-fen Huang for kindly providing the CEFR ratings of the 45 CzE speakers and 20 GerE speakers.

[22] Czech and German components of the *Louvain International Database of Spoken English Interlanguage* (Gilquin et al. 2010).

smaller L2 vocabularies. However, they only propose this hypothesis for learners at an early point in L2 immersion, and larger L2 vocabulary at a later point forces the learner to settle on a possibly accented L2 phonology. Different measures are proposed to investigate the lexical sophistication of spoken and written learner data, e.g. lexical complexity, lexical density, type-token ratio (TTR), lexical bundles, and vocabulary frequency band (Biber 1988; Biber 2006; Staples et al. 2013, 2017; Staples and Reppen 2016; Yan and Staples 2019). Examining lexical development as the L2 proficiency levels increase, lexical diversity, density and complexity measures, are generally found to be higher in writing than in speech, lower for learners than NSs, and higher for advanced learners. However, in Ishikawa's (2015) study lexical density remained unchanged in spoken and written language throughout the different proficiency levels, while lexical diversity decreased and then increased, and lexical complexity was identified as the only measure that steadily increased as a function of increasing L2 proficiency.

All in all, lexical knowledge has been recognized as an important part of language proficiency and several studies found lexical diversity to be a strong predictor of learners' proficiency levels (e.g. Yu 2009; Kyle and Crossley 2015). Therefore, lexical complexity will also be investigated in the present study. A vocabulary-phonology link might give an indication of the learners' proficiency levels, across developmental stages and speaking tasks (monologues vs. dialogues). Previous research showed that TTR exhibits the strongest correlation with language ratings of essays (Staples and Reppen 2016: 27). Similarly, a strong link between fluency and prosody has been claimed by researchers, such as Saito (2015) and Trofimovich and Baker (2006), who found that fluency and prosody improve as a function of LoR in an English-speaking country. Similarly, Lintunen et al. (2015) also found that their Finnish L2 learners of English approached NS fluency level with increasing proficiency and they showed that the intonation phrase (tone unit in their terms) can be used as a base unit in measuring L2 fluency development (see also (Gut 2012: 15) who also defined MLR as the average number of syllables per intonation phrase). However, prosody may require more L2 experience, i.e. a higher L2 proficiency, than fluency (cf. Saito 2018: 286). Thus, the longer the stays abroad and the higher the fluency of the speakers, we can assume the higher the proficiency in L2 prosody in terms of prosodic development. Therefore, the fluency ratings for the individual learners from Götz (2013) and Gráf (2015) will be considered as well, since they are based on the exact same learners as in the present data set, i.e. LINDSEI-GE and -CZ, respectively.

1.5.4 Perception vs. production

As will be shown in Section 2 of the present study, learners of English from different L1 backgrounds seem to exhibit different degrees of difficulties in their production of prosodic features. Even though this study is only concerned with L2 production of prosodic features, insights into the perceptive skills could have some explanatory power. For instance, the learners' deviant prosodic productions could be explained by their perceptive skills. Therefore, it is important to briefly consider the production-perception link in the L2 acquisition process of prosody.

While current L2 speech learning theories (Best 1995; Flege 1995) were primarily developed for segmental learning, they point towards an important interaction between perception and production in the successful acquisition of L2 phonology. Some studies have found support for the claim that correct perception is a precondition to correct production in adult prosody production abilities (perception-first: Cutler and Swinney 1987; Flege et al. 1997a; Chen 2010; Lentz and Chen 2015; Colantoni et al. 2016b), while others found that production precedes perception (e.g. production-first: Sheldon and Strange 1982; Bundgaard-Nielsen et al. 2012; Lentz and Chen 2015; Puga et al. 2018). In line with the perception-first view point, the perceptual intonation training of Dutch learners of English yielded a statistically significant improvement in the L2 production of intonation patterns in a study conducted by De Bot and Mailfert (1982). According to 't Hart and Collier (1975, as cited from Chun 2002: 91) a 12-minute tape about the perception of intonation lead to a significant improvement of L2 productions of Dutch learners of English. Even though such a direct and immediate link between perception and production is usually not found in L2 learning, a tight and complex link between the two, which is modulated by various phonetic, phonological, and social factors, has been postulated in previous research (Sheldon and Strange 1982; Best 1995; Flege 1995; Flege et al. 1997a; Grosvald 2009; Lin 2011; perception-production in the L1 see Tyler 2012; Lentz and Chen 2015; Speeter Beddor 2015). Based on segmental features, Kartushina and Frauenfelder (2014) did not find a close correlation between perception and production, which may suggest some kind of dissociation between the two levels. Despite contradictory results, most researchers recognize that due to the interconnected nature of perception and production abilities that L2 learners simultaneously improve both abilities over time (cf. Bundgaard-Nielsen et al. 2011, 2012). Some studies suggest that L1 transfer and difficulties that L2 learners experience are more evident in L2 productions than in perception (e.g. Ortega-Llebaria and Colantoni 2014), since there are many more opportunities in production for the L1 to shine through, which would further support the claim that perception precedes production. Depending on the similarities and differences between the L1 and L2, L2 perception may be hindered in L2 production due to influences from the L1, as has been shown in different speech learning models in Section 1.4.3. L2 learners may perceive the L2 through L1 categories (e.g. Grabe et al. 2003) and this can have an effect on their L2 production. Studies on the perception of prosodic features additionally show that learners of English lack skills in the perception of prosody (e.g. Mok et al. 2016; Puga et al. 2017). In a study including different sentence types such as statements, questions, etc. in a read short story, Puga et al. (2017) found that German learners of English performed similar to British controls for sentence types, such as statements and *yes*/*no*-questions, but they performed worse in the case of open and closed tag questions and the expression of sarcasm (see similar results with SpE learners in Ramírez Verdugo and Romero-Trillo 2005: 162). Strikingly similar results were previously found by Mok et al. (2016) with Hong Kong English speakers. While NSs of English produce tag questions with falling and rising intonation, and a combination of these, NNSs tend to produce more rising tones and overgeneralize them to all contexts. Ramírez Verdugo and Romero-Trillo (2005: 163) explain that NNSs have pragmatic difficulties by interpreting every syntactic subject-verb inversion as a polar question and ignoring the linguistic and situational context. These findings suggest that L1 influence could be one possible explanation for the results (Ramírez Verdugo and Romero-Trillo 2005: 163-64; Puga et al. 2017), since for instance tag questions in Spanish are primarily used to demand confirmation and are used with rising intonation (Quilis 1993: 451-452). Puga et al. (2018), who analyzed the production-perception link, found the production of these same learners to be more native-like than their perceptive skills. While their performance in the perception task showed similarities with the native control group, substantial differences were found in production, especially in particular sentence types (e.g. closed tag questions and checking questions). Surprisingly, the learners also performed significantly better than the NSs in two contexts (yes/no and statement questions). Puga et al. (2018) discuss the possibility that the learners might tend to adopt standard BrE intonation tonal choices, while the native control group might follow a more recent model that is the result of ongoing language change not yet reflected in ELT materials (e.g. text books). The reader is referred to Lindsey (2019) for a discussion of an outdated BrE pronunciation model, in which he states that BrE is becoming more similar to AmE pronunciation on the segmental and intonational level. In sum, previous research implies an intricate relationship of production and perception of L2 prosody. While learners may have difficulties in perceiving and producing for instance tag questions, because they rely more on syntactic structure instead of using appropriate pragmatic functions and tonal choice for a given context, the production of statement intonation may not pose any problems for learners of English. Whether statements are produced in a native-like manner without influence form the L1 by learners with different L1 backgrounds, will be investigated in the present study.

1.5.5 Stay abroad and learning context

It has been argued that the improvement in L2 prosody is rather "gradual and slow in nature" (Saito 2018: 286). However, the LoR in an English-speaking country generally influences improvement, as a number of studies report (e.g. Flege and Fletcher 1992; Trofimovich and Baker 2006, with Korean learners of English; and other L1-L2 contexts: Ullakonoja 2007, with Finish learners of Russian; Saito 2015, with Japanese learners of English). Some studies even report on near-native-like L2 speakers of English after a stay abroad or permanent residence (e.g. Mennen et al. 2014); others report on sobering results of longitudinal studies with the LoR variable (no improvement: Spanish learners of English after 3-year stay abroad: Ramírez Verdugo 2003: 662; Gut 2009, 2017: six to nine months), while others have mixed results (Henriksen et al. 2010: seven-week stay; Derwing and Munro 2013: improvement for Slavic but not for Mandarin learners of English after a seven-year period). Götz and Mukherjee (2018: 54, 58) found a significant positive correlation between most of their analyzed variables (discourse markers and smallwords, grammatical accuracy, and lexical development) in LINDSEI-GE and the stay abroad variable. However, they observed an increase for article errors after a long stay abroad (ibid.: 59). Therefore, whether L2 proficiency improves depends (among other factors; see Section 1.5.3) on the linguistic features under investigation (see also Trofimovich and Baker 2006) and the LoR. Yet, not all learners seem to improve their L2 proficiency after a stay abroad. For instance, Birdsong (1999: 15) observed a 10-15% success rate of native-like L2 attainment of learners who lived in a favorable L2 environment for a substantial amount of time (three to five years). Previous research suggests that L2 speech continues to change, develop, and improve over a long period of time (e.g. Flege et al. 1997a; Saito 2015; six to ten years cf. Flege 2016) and that it becomes stable, provided that the learners predominantly communicate in their L2, which entails a reduction of L1 use. It is a common assumption that a LoR of five to ten years or more does not significantly influence L2 performance (e.g. Flege 1988; Abrahamsson and

Hyltenstam 2008, 2009; DeKeyser et al. 2010). Previous research has found that accents generally tend to stabilize after 1 year of residence in an English-speaking country, provided that there is no explicit pronunciation instruction (e.g. Flege 1988). Derwing and Munro (2013) found that Slavic learners of English improved their comprehensibility, fluency, and accentedness after two years of residency. However, several studies have shown that L2 performance does become more native-like after ten years or more of residence (e.g. after six years: Saito 2013; LoR only has an affect if learners receive a substantial amount of NS input: Flege et al. 1995; Flege 2009). However, the accentedness did not improve beyond the two-year mark after another analysis at the seven-year mark.

Not only the length of a stay abroad (shorter and longer stays) plays a significant role in the acquisition of prosody for some individuals but also discontinued L1 use might lead to a significantly reduced foreign accent, as compared to speakers who continue using their L1 in an L2 environment (e.g. Thompson 1991; Flege et al. 1997b). In contrast to that, other researchers argue that L1 maintenance does not hamper L2 native-likeness (e.g. Bylund et al. 2012).

Previous studies also suggest that fluency might develop more quickly and earlier with increased LoR (Derwing et al. 2009, especially for Slavic learners of English compared to Mandarin learners), prosodic features, such as stress and intonation, may require a greater deal of L2 exposure and experience in order for the learner to internalize native-like usage (Trofimovich and Baker 2006, with Korean learners of English). This is an interesting finding for the present study, as will be shown in Section 2.6, since fluency is very frequently correlated with intonational phrasing, and it might, therefore, be possible for learners to simultaneously improve some aspects of prosody (i.e. intonational phrasing), while other prosodic features might need more time (e.g. stress and peak alignment).

What might be more impactful than the LoR variable are the learning context and the intensity of the interactions with NSs (cf. Dietrich et al. 1995: 277). Rarely do learner corpus studies define the learning context, proficiency levels, as well as the advancedness of learners in a satisfactory manner (e.g. Ortega and Byrnes 2008; Callies 2009). It is only logical that learning does not take place if the learner lacks the willingness to communicate with NSs in the L2 environment (Flege and MacKay 2004; Derwing and Munro 2013). Nevertheless, an interaction can be assumed between the length of a stay abroad and the likelihood of opportunities for interaction (cf. Götz and Mukherjee 2018: 49). Equally, the purpose of the stay abroad may influence the acquisition of prosodic features, such as stays abroad as aupairs, work purposes in the case of immigrants (e.g. Mennen et al. 2014), or semester abroad purposes in a university context. Gut's (2017) study offers some particularly interesting findings for the present study. In her study, participants improved their L2 phonology after a stay-abroad program in story retellings and in free speech, more so than in read speech. She concludes that "[a] real-life communicative context can especially support the development of spontaneous speech and narrative abilities" (Gut 2017: 218). Thus, the learning context seems to have an influence on the development of phonological features most apparent in spontaneous speech. The learning context of the learners in the present study is limited to the LoR, the country they resided in, the length of English instruction in school and university, and other foreign languages they have learned.

1.5.6 Prosodic training

As has been described in Section 1.2, prosodic instruction is generally not part of most learners' learning context in EFL classrooms. However, positive effects for L2 phonological development have been found through prosodic instruction. General formal English language instruction in school and university has been found to have an influence on L2 phonology (e.g. Flege and Fletcher 1992), and positive effects were also found for long-term phonetic training (Moyer 1999; Gut 2009), and audio-visual instruction (De Bot and Mailfert 1982; De Bot 1983; Michas and Berry 1994; Levis and Pickering 2004; Hardison 2004, 2005; Pincus 2014). In particular, learners of English who received instruction in rhythm, stress, and intonation have been shown to significantly improve their overall accent (Derwing et al. 1998; Missaglia 1999; Derwing and Rossiter 2003; Smorenburg et al. 2015; Saito and Saito 2016). Missaglia (1999) tested prosody- and segment-centered instruction (20 hours of training for each group) in Italian learners of German and found that prosody-centered instruction had a larger effect on a global accent-improvement, i.e. segmental and suprasegmental features. In Derwing and Rossiter's (2003) study positive effects after a 12-week ESL course were found on comprehensibility and fluency through judgments by native raters on the learners' spontaneous narratives elicited from a cartoon story, only in the prosodic (stress, intonation, rhythm, and speech rate) type of instruction and not segmental instruction (no change was observed for the segmental instruction). Interestingly, Derwing and Rossiter (2003: 12-13) concluded that segmental training might focus too much on form and, as a consequence, meaning and context appear to suffer. Prosodic training involves the training of different contextual situations and might ultimately improve the learners' overall accent (ibid.). In a study with German learners of AmE, Moyer (1999) found a closer-to-native performance was indicated by both overt segmental and suprasegmental training and feedback. Other studies report that segmental pronunciation training alone has no effect on a native-like performance or improved foreign accent (Suter 1976; Purcell and Suter 1980; Thompson 1991; Flege et al. 1995; Derwing et al. 1998). Gut (2009: 247), however, reported a greater influence of a six-month long prosodic training on the acquisition of intonational patterns (i.e. stress placement and increase of pitch movement and complex nuclei) than the stay abroad variable although more advanced learners (proficiency based on accent ratings) profited more from prosodic training. She explains this result as follows: "Possibly, less advanced learners have less capacity to take in and put into practice prosodic rules because they are more occupied with the more 'basic' tasks of word retrieval and grammatical aspects of speech production" (Gut 2009: 281). Saito and Saito (2016) found empirical evidence that prosodic training leads to an overall phonological improvement even with Japanese beginner-level EFL learners with limited L2 conversational experience. Previous research (e.g. Gut 2009; Henriksen et al. 2010) revealed large interspeaker variability and L1 transfer alone cannot account for prosodic variation in L2 speech and speaking styles and learning contexts have to be considered. In sum, it is likely that a combination and intricate interplay between the various variables discussed so far, such as LoR, AoA, and prosodic training, contribute to native-like L2 attainment.

1.5.7 Corpora, SLA, and prosody

In the 1960s, the first linguistic descriptions of L2 productions emerged, and explanatory theories of second language acquisition (see Section 1.4) and factors (see Section 1.5) that condition this process have been proposed (cf. Gut 2014: 286). Since the 60s until today, a variety of research methods has been applied in SLA and since the 1980s, spoken corpora including productions of language learners, i.e. learner corpora, have been collected by various researchers (cf. Granger 1998, 2004; Granger et al. 2002; Pravec 2002; Myles 2005; Gilquin et al. 2010; Ward and Gallardo 2015; Granger et al. 2016; Trouvain et al. 2017). For instance, more recently collected corpora that can be mentioned are the *Louvain International Database of Spoken English Interlanguage* (LINDSEI) (Gilquin et al. 2010), and the *Louvain Corpus of Native English Conversation* (LOCNEC) (De Cock 2004). However, none of these learner corpora contain phonetic or prosodic transcriptions. One exception is the *Learning Prosody in a Foreign Language* (LeaP) corpus (Milde and Gut 2002; Gut 2009, 2010, 2012), which is the first freely-available corpus to be compiled specifically for the study of non-native and native productive prosody. However, corpus-based research on L2 prosody is still very rare: “So far, however, corpus linguistics and second language research have mainly coexisted side by side, and they are only just beginning to join forces” (Gut 2014: 300). While the interest in a combined corpus-based approach to non-native speech is gaining popularity, as recent volumes evidence (e.g. Friginal et al. 2017; Le Bruyn and Paquot 2020; Tracy-Ventura and Paquot 2021), a focus on prosody is still outside the mainstream. Corpus-based CIA analyses of prosody that tested explanatory SLA theories against corpus data are practically non-existent for intonation languages; a few exceptions are Gut (2009), Tortel and Hirst (2010), Tortel (2013), and Ward and Gallardo (2017). In Gut (2009), however, a clear distinction between L1s is not always made (only for f0 range in Spanish, Russian, and German speakers). Table 3 in Section 2.3, which gives an overview of L2 prosodic studies of Czech, German, and Spanish learners of English, shows that a corpus-based approach is rarely adopted. Even concerning some studies that are termed “corpus-based” (e.g. Gut 2017; Puga 2019), one could argue that, due the small size in the number of participants, they should not be classified as corpus-based in a strict sense. See Section 2.3 for a more extensive overview of current and past L2 prosody research methods.

Previous studies on L2 prosody from Czech, German, and Spanish L1 speakers have identified a range of deviances and possible explanations for these, i.e. L1 transfer and developmental errors. A more detailed overview of the features investigated and the findings of these studies are presented in Section 2. However, there is a range of limitations in comparability of previous findings that have to be mentioned. For instance, studies on L2 prosody mostly base their findings on experimental data, and in some cases extremely controlled laboratory data is used. Furthermore, a limited number of participants is typical for intonation studies, while studies are generally restricted in one way or the other, i.e. analysis of only one speaking style (read speech) or one isolated aspect of non-native speech (mostly nuclear stress). In addition, extralinguistic variables, such as gender or learning context, are usually not taken into account or are often circumvented by including only female or male speakers in the analysis. Complicating matters, a high diversity of proficiency levels (ranging from beginners to highly advanced learners) and different theoretical views on intonation including different tonal inventories and annotation systems used makes the comparability of data in L2 prosodic studies extremely difficult. Therefore, a corpus-based approach with sufficient data could compensate for the shortcomings of previous studies by including more quantitative findings on L2 prosody.

Corpora make quantitative data analyses possible and, therefore, make it possible to quickly and easily find errors that learners frequently make or to identify deviant usage of certain linguistic structures. Even unsuspected linguistic features may be uncovered. The prosodic features analyzed will not be treated as 'errors' but rather as deviations from the target norm. To find possible explanations of prosodic deviances in non-native speech, a combined approach of CIA, CA, SLA, and corpus linguistics is essential. In light of this, the present study can best be described as a linguistic and quantitative corpus-based and -driven explorative analysis of prosodic features of native and non-native speakers of English, such as intonation and intonational phrasing in narrative and informal interview contexts. Thereby, I apply a CIA (Granger 1996, 2015) and investigate prosodic features of Czech, German, and Spanish learners of English. In doing so, the present study adds to the scarce corpus-based SLA analyses. The language-based models and their predictions discussed in Section 1.4 will also be put to the test in the present study, the LILt model specifically.

1.6 Summary

The first section of the present thesis started with the aims and research questions as well as definitions of prosody. This was followed by a brief excursion into the English language learning situation in the Czech Republic, Germany, and Spain. ELT as well as prosody instruction have been found to be comparable in the three countries. Then the native-non-native continuum was discussed in Section 1.3 and the NS norm was determined to be a useful benchmark for the purpose of the present study. Additionally, the previous sections reviewed major SLA theories and their main predictions for L2 phonetic and phonological acquisition of an L2. This included a brief discussion of L2 speech learning models (cf. Section 1.4), which focused on the LILt working model, put forward by Mennen (2015), whose predictions will be tested in the present study, according to its four intonational dimensions (systemic, realizational, distributional, and functional). Another theory that will be tested in the present study is the CAT. The previous section has shown that accommodation to native as well as non-native speech is conditioned by social attitudes (Bresnahan et al. 2002; Babel 2010; Dragojevic and Giles 2016), the experience of NSs with (non-)native dialects, if the foreign accent is within the speaker's articulatory repertoire (Kim et al. 2011), and is subject to individual variation (Lewandowski and Nygaard 2018). In the present study CAT will be tested in terms of patterning of f0 and entrainment by all the available sociolinguistic variables.

All social-psychological and contextual variables (task conditions and interviewer role) discussed in Section 1.5 contribute individually to the successful acquisition of an L2, which ultimately can lead to native-like proficiency in the L2, either the pronunciation of a learner on the whole, or prosody in particular. Most extralinguistic factors, such as age, gender, LoR, motivation, and length and type of instruction, have diverging results and influence the L2 production to a different degree. While some studies show positive correlations, some show none, or mixed effects. This is highly dependent on the participants as well as the context of the study. While most other factors are quite difficult to measure, and whether one factor is particularly 'responsible' for successful ultimate attainment is difficult to prove. It depends on how one defines success and which linguistic level one refers to. While factors such as a learner's intelligence and language learning aptitude may give the learner a head start and speed up the initial learning process, they are not a prerequisite for successful ultimate attainment. However, one of the most dynamic psychological factors, motivation, can be regarded as the most important factor contributing to successful L2 acquisition. Someone can be exceptionally intelligent and talented, but if that person is not motivated, then the language will never be fully acquired. All of these factors reviewed in Section 1.5 are intertwined and may influence each other, e.g. a learner's sex is likely to interact with other variables that determine successful L2 attainment, such as higher motivation, more positive attitudes toward L2 speakers, and different learning styles, all of which have been observed in female learners. If possible these extralingustic factors should therefore not be investigated in isolation, and as many confounding variables as possible should be included in the analysis, since the acquisition of a foreign accent is a rather complex construct.

The sociolinguistic variables in the present study are restricted to dialect (i.e. variety of English), sex, age, stay abroad, language proficiency (different measures), learning context, and other extralinguistic interview situation variables. The perception of prosodic features by the learners themselves cannot be determined in the present study. Furthermore, ethnic variety and social-class cannot be analyzed, because all speakers in the present study are Caucasians from a similar socioeconomic background. In sum, the present study adds to the growing corpus-based research discussed in Section 1.5.7 and adopts a quantitative mixed-method and multifactorial approach.

2 Native and non-native prosody

The goal of Section 2 is to give an overview of the literature that informs and motivates the research questions of this dissertation project. Specifically, previous studies on native and non-native prosody, theoretical concepts, and various methodological approaches to the study of L2 prosody will be summarized and discussed, including different aspects of how L2 speakers of English from different L1 backgrounds *produce* and *acquire* intonational patterns of English. In doing so, I will look closely at the areas in which learners of English deviate from native English prosody. Prior to the discussion of the different forms and functions of prosody with regard to native and non-native speech, it is relevant to consider different models of intonational structure, on which the previous studies of prosody (form and function) were based (see in Sections 2.5-2.8). The two main models that will be outlined in Section 2.1 are the *autosegmental-metrical* (henceforth AM) *approach* (American tradition) and the *contour-based British tradition* (British School). Additionally, relevant L2 speech models that include predictions on L2 prosody will be briefly discussed. An overview on how the different L1s (German, Czech, and Spanish) in this study differ from English prosody will be described in Section 2.2. Section 2.3 will then identify problem areas for Czech, German, and Spanish learners of English in terms of prosody production. Sections 2.5-2.8 review previous research on the four prosodic features selected for the investigation of native and non-native prosody, i.e. tones and tunes (Section 2.5), intonational phrasing (Section 2.6), f0 range (Section 2.7), and uptalk (Section 2.8).

2.1 Models of intonational structure: The autosegmental-metrical approach vs. the contour-based approach

The methodological framework of autosegmental-metrical phonology and its instrumental paradigm (ToBI; described in Section 3.4), was selected as basis for the phonological analysis of the present study. Why it has been selected and why it is the best choice for the present study will be explained in the present section. Many different models of intonational structure have been proposed over the last five decades for the purpose of describing intonation patterns. As outlined above, the two models of intonational structure, which are the most widely known and employed, are the AM approach associated with the American tradition (Pierrehumbert 1980; Beckman and Pierrehumbert 1986) and the *contour-based* approach of the British School tradition (e.g. Palmer 1922; Kingdon 1958; Halliday 1963, 1967, 1970; O'Connor and Arnold 1973; Tench 1996, 2015).[23] Both models are based on fundamentally different assumptions and make use of different notation systems. Table 2 summarizes the differences and similarities between the two approaches. For a more detailed overview of the AM approach, the reader is referred to Beckman and Pierrehumbert (1986) and Ladd (1996; 2008). For a detailed overview of prosodic models, see Anufryk (2012), who reviews a myriad of different prosodic models (e.g. IPO by Öhman 1967; Fujisaki Model by Fujisaki

[23] The reader is referred to Pickering (2010) for an overview of the history of approaches to intonational analysis and Levis (2005) for a detailed comparison of the British and American traditions of intonational analysis.

1983; PaIntE[24] by Möhler 1998, 2001; Tilt model by Taylor 2000; INTSINT by Hirst and di Cristo 1998, etc.).

Table 2. Overview of differences between the AM approach and the British School of intonational modeling

<table>
<tr><th></th><th>AM approach</th><th>Contour-based approach</th></tr>
<tr><td>Definitions</td><td>Consists of an autosegmental tier representing intonation's melodic part (association of tune with text) and a metrical structure representing prominence and phrasing</td><td>Contour-based approach of intonational analysis</td></tr>
<tr><td>Purpose</td><td>American generative tradition</td><td>British functionalist and pedagogical traditions</td></tr>
<tr><td>Primitives (basic units)</td><td>• Pitch levels High (H) and Low (L)
• Smallest elements are tones (H or L)
• No necessary role ascribed in pitch description to the syllable</td><td>• Rise or fall are the primitives
• Smallest elements are syllables</td></tr>
<tr><td>Pitch movement</td><td>• Pitch movement is usually called intonation</td><td>• Pitch movement is usually called the tone</td></tr>
<tr><td>Intonational analysis</td><td>• Pitch contours as sequences of pitch levels (H & L targets & combinations)
• The actual contour results from an interpolation of these underlying tones
• Recognizes and separates a phonological (set of H and L tones) and a phonetic component (scaling and alignment) in intonation
• No theoretical distinction between 'pre-nuclear' & 'nuclear' accents made
• Intonational analysis relies on a combination of computer-assisted instrumental and auditory techniques</td><td>• 'Contour' analysis of dynamic pitch configurations or pitch movements (i.e. falling or rising pitch movements)
• Posits a finite inventory of tunes for a given language (e.g. O'Connor and Arnold 1961)
• Most important contour: nuclear tone
• Auditory method, rather impressionistic</td></tr>
<tr><td>Transcription convention</td><td>• Targets specify only specific points in the f0 contour, represented phonologically as 'tones' of three types: pitch accents (*), phrase accents (-), and boundary Tones (%)
• Pitch accents associate with metrically strong (prominent) syllables
• Edge tones associate with structural positions, such as phrase boundaries (phrase accents) and utterances (boundary tones)
• IPs consist of at least one pitch accent and an edge tone</td><td>• Intonation represented in detailed interlinear transcriptions or tadpole diagrams depicting the properties of each syllable in terms of accentedness, pitch height, and pitch movement by the use of dots
• It has been called interlinear, since the transcription is placed between two lines indicating the upper and lower limit of a speaker's f0 range</td></tr>
<tr><td>Phrasing</td><td>Distinction made between smaller ips & larger IPs</td><td>The basic units of the intonational structure are the tone units (TU). Minor and major tone units (MU) can be distinguished (Trim 1959)</td></tr>
<tr><td>Key terms</td><td>Prosodic hierarchy (Pierrhumbert 1980)</td><td>Tonicity, tonality, and tone (Halliday 1963; 1967; 1970)</td></tr>
</table>

As Levis (2005: 342) contends, the BrE and AmE traditions of intonational analysis have a many characteristics in common, i.e.

[24] Parametric INTonation Events.

> Both traditions thought that contours were the meaningful elements of intonation, both asserted that the contours could be described as made up of primitive elements, both thought the meanings of individual contours were abstract and could not be tied to particular grammatical structures, and both asserted a far greater number of meaningful distinctions than are now taught in any textbook.

However, there are many differences as well. One can observe from Table 2 that one of the main differences between the two approaches is that *dynamic pitch movements* or *contours* (i.e. falling or rising pitch movements) are the basis of intonational analysis in the contour-based approach, whereas the AM approach takes *sequences* of pitch *levels* as the basis. How these f0 patterns are represented in the two intonational models is explained in Warren (2016: 13) and Pierrehumbert and Hirschberg (1990: 281). It is important to understand the different intonational models since, for instance, the differences between different AM analyses might very often be "of a theoretical rather than a typological nature" (Grice et al. 2005: 56). As summarized in Table 2, the AM approach consists of a two-level approach to intonation: 1) an autosegmental tier representing intonation's melodic part (association of the tune with the text) and 2) a metrical structure representing prominence and phrasing (cf. Liberman and Prince 1977). A great advantage of the AM approach is that it recognizes both a phonological and a phonetic component in intonation and it separates the *phonological representation* from its *phonetic implementation*. For instance, a phonological influence in non-native speech would be the replacement of rises with falls in sentence types where the opposite tone would be expected (e.g. Gut 2009; Puga 2019), since it has implications on the comprehension of the utterance. An example of a phonetic influence would be the narrower f0 range often found in L2 prosodic studies (e.g. Mennen et al. 2014; Volín et al. 2015), which usually does not have implications on the comprehension of utterances. Intonational tunes in the AM approach are analyzed as sequences of the primitives high (H) and low (L) which form the phonological representation of the tune. The scaling (i.e. the f0 value) and the temporal alignment with the segmental strings form the phonetic implementation. The actual tune of these pitch levels is interpreted by interpolations between the individual pitch levels. That is, if there is a H* H-L% tune, the AM model assumes that between the first H* and the second H- there has to be some kind of a low in-between. Thus, the movement to and from these targets is seen as phonetic interpolation. The British School takes an opposite view where pitch movements are seen as the most important phonological categories of intonation. Figure 1 shows the structure of intonation according to Beckman and Pierrehumbert's (1986) revised version of the AM approach.

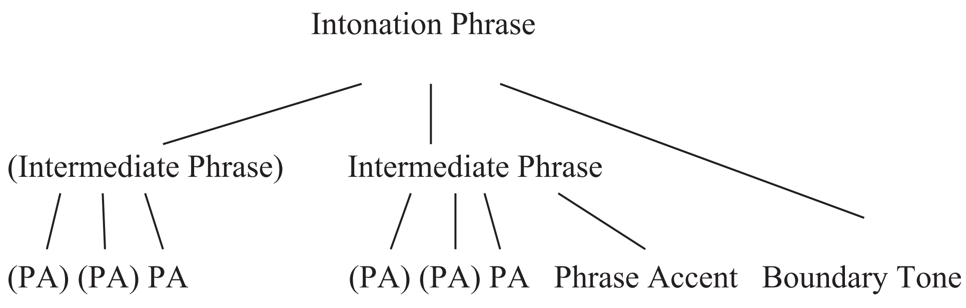


Figure 1. The structure of intonation according to Beckman and Pierrehumbert (1986) (see also Gut 2009: 199)

As can be seen from Figure 1, tones are the smallest basic units in the structure of intonation according to the AM approach. The highest level is the intonation phrase (IP), which can consist of at least one or more optional intermediate phrases (ip). Thus, there are some ips that finish within an intonational phrase and one that finishes at the same point as an intonation phrase. Each ip, according to the AM model, consists of at least one or more *pitch accents* that are associated with stressed (or prominent) syllables. Besides pitch accents, there are also what I will call from now on *edge tones*, which is used as a general term for any tone that appears at the periphery of ips or IPs (cf. Ladd 2008: 47). There are two different types of edge tones, i.e. *phrase accents* occurring after the last pitch accent at the periphery of ips and *boundary tones* occurring at the beginning or periphery of IPs (cf. Pierrehumbert 1980). These edge tones can occur on stressed or unstressed syllables. Since, according to the Strict Layer Hypothesis (SLH), the end of an IP is at the same time the end of and ip, there are four combinations of possible edge tones at the end of IPs, i.e. L-L%, H-L%, L-H% and H-H% (cf. Warren 2016: 11-12). In the present study I follow Mennen's (2007: 173) recommendation in implementing both a phonetic and phonological level in the analysis of prosody, in order to be able to make predictions on the difficulty learners might have in producing L2 prosody. Therefore, I will focus on all possible combinations of H and L tone productions (phonological) in declarative utterances and their f0 range (phonetic). There are three main reasons for the selection of declaratives only, 1) statements are the most frequent sentence-type in the LINDSEI and LOCNEC corpora, 2) previous research found phonetic and phonological levels of prosody to differ by sentence-type (Colantoni et al. 2015), 3) to make the data more comparable and reducing the complexity of the analysis, i.e. even statements can have variable phonetic and phonological variability depending on the context. An intonational phrase domain is also necessary in order to compare the spontaneous utterances of differing lengths with each other. The present study is, thus, original in examining tonal choices and f0 range by speakers from different L1 groups and native speech.

If this is compared to the British-style analysis (e.g. Halliday 1963, 1967, 1970; Crystal 1969; O'Connor and Arnold 1973; Tench 1996, 2015), which treats intonation in terms of dynamic pitch contours, the most important contour is the 'nuclear tone' (Halliday's 'tonic') on the utterance's most prominent syllable. According to Beckman and Pierrehumbert's revised understanding of the phrase accent, the 'nuclear accent' can be seen as the last accent of the ip (cf. Ladd 2008: 90). However, in the AM approach no distinction is made between pre-nuclear and nuclear tones. Intonational phrasing (including IPs and ips) in the AM approach can be compared to the 'major' and 'minor tone units' respectively in the British school (Williams 1996a, b). In the British school tradition the analysis of intonational contours is mainly conducted auditorily and impressionistically, while the AM approach relies on a combination of auditory analysis and computer-assisted instrumental techniques (cf. Gut et al. 2007: 14).[25] In British School analyses a dot is assigned to every syllable, stressed syllables appearing in larger dots than unstressed ones. The pitch movement on prominent syllables is represented by lines. For this reason, the form of this notation has been called *tadpole notation*. The speakers' f0 range is represented by the two lines (hence the term *interlinear transcription*). The AM notation system, on the other hand, treats tunes as sequences of pitch levels and interpolations between these levels by using two primitives (i.e.

[25] The different notation systems of the two approaches are visualized and explained in Jilka (2007: 80).

H and L) in three different positions (*, -, %). The AM notations are explained in more detail in Section 3.5.

While there is no disagreement over the fact that intonation carries meaning, there is disagreement on the parts of intonation that carry a particular meaning. According to Pierrehumbert and Hirschberg (1990), all individual components of a tune (i.e. pitch accents, phrase accents, and boundary tones) contribute together to a composite meaning of a tune. An opposing view to this is that the tunes themselves (rise, fall, and fall-rise) are the meaningful elements associated with abstract meanings (cf. Gussenhoven 2004: 316-320; Ladd 2008: 147-156). In my analysis in the present study I adhere to the former compositional account of intonational meaning, because this is the viewpoint adopted by the AM approach.

The present study adopts the AM model. Several intonationists have argued for an analysis of levels over one by configurations (e.g. Ladd 2008; Arvaniti 2017) and more L2 prosody research seems to be conducted with an AM-based approach (see Table 3 in Section 2.3). Besides the advantage of a direct comparison to previous research that also adopted the AM approach to the analysis of L2 prosody, the choice of the AM model is also explained by its capability of accounting for phonological variation in intonation and the availability of ToBI labeling systems for English, Czech, German, and Spanish. However, no ToBI labeling system for the different non-native varieties of English exists, which is why the present study opted for a combined inventory of the four languages. The exact labeling approach is described in sections 3.5 and 3.7.1.1. In the present study, in accordance with the AM approach, the primitives high (H) and low (L) are seen as pitch levels (see the levels-vs-configuration debate, cf. Ladd 2008) that are inherently static units devoid of any information on their directionality. Just as the AM approach proposes, no distinction between the status of nuclear and pre-nuclear pitch accents is made. According to the AM approach of intonational modeling, syllables, which are part of the segmental as well as suprasegmental study (e.g. in speech rhythm), are not ascribed any necessary role in pitch description. The present study follows the practice of the AM approach and does not analyze syllables, vowels or consonants. In this thesis, I chose the same approach as Cruttenden (1997: 13), where he uses the term *stress* to refer to *prominence*, no matter how such prominence is achieved. The term *accent* is limited to prominences where pitch is involved (hence it is equivalent to *pitch accent*). In the present study I investigate word-stress and sentence-stress only insofar as they are a prerequisite for intonation (cf. Cruttenden 1997: 14). In doing so, I follow the AM approach by distinguishing between stress and accent. Thus, stress is regarded as a *potential* accent and accent as the realized stress (resulting in perceived prominence) when a word is produced in an utterance (cf. Gut et al. 2007: 6). The theory of Pierrehumbert and Hirschberg's (1990) compositional approach to intonational meaning as well as the meaning of individual tones will be tested in the present study, and it will be determined whether the tones and tunes in the present data carry the same meanings as described by Pierrehumbert and Hirschberg (1990). For an alternative to theory-driven analyses of prosody, the reader is referred to Ward and Gallardo (2017) and Ward (2015) for an automatic PCA-based (Principle Component Analysis) method for the analysis of prosodic constructions. Even though the AM approach was selected for the present study, when previous research is cited, their terminology stemming from a different model of intonational structure may be used as well.

2.2 Comparing German, Czech, and Spanish prosody to English prosody

Several studies have reported deviances of L2 prosody from the native or target norm (e.g. Ramírez Verdugo 2002, with Spanish learners of English; Gut 2007, 2009, with German learners of English; Volín et al. 2015, with Czech learners of English), which are often attributed to L1 influence and transfer and have identified universal patterns as well as developmental processes of intonational acquisition. Whether L1 influence can be considered to be one of the main factors explaining most of these deviances (see Section 2.3 for an overview) or if additional explanations have to be taken into account, it is first necessary to review how English, Czech, German, and Spanish differ generally in terms of prosody. This section provides a brief contrastive overview of Czech, German, and Spanish in comparison to English prosody.

Czech and English are both intonation languages. While English has a rather stress-timed rhythm, Czech has a rather syllable-timed rhythm,[26] but can also be described as an intermediate mixed type between stress and syllable-timed rhythm (cf. Barry and Andreeva 2001). Czech has relatively free word order, whereas English has a fixed word order; yet, both languages have a tendency to keep the nucleus in final position (cf. Duběda and Mády 2010). Czech may make use of grammatical inflection or word order when expressing contrastivity and signaling major information instead of relying on intonation, as would usually be the case in English (cf. Rogerson-Revell 2011). Additionally, word stress has been described to differ in Czech and English, since Czech has a fixed position, i.e. the first syllable receives word stress and it usually does not involve an increase in f0,[27] intensity, or duration (cf. Weingartová et al. 2014a: 236). In English lexical stress is contrastive and is characterized by longer durations, a flatter spectral emphasis in decibels (dBs), and higher f0 range (Eriksson and Heldner 2015). Studies on sentence-level f0 range have shown that Czech has a narrower f0 range than English (cf. Volín et al. 2015; for more discussion on f0 range differences, see Section 2.7.2). Recently, researchers have dedicated much research to describing the differences between Czech and English prosody, as well as Czech English and English Czech L2 prosody (e.g. Volín et al. 2015, 2017).

Since German and English are both Germanic languages, there is a small genetic and typological distance between these two languages (e.g. Esser 1978). For instance, German and English are both described as intonation languages and both are located on the same end of rhythmic continuum, i.e. stress-timed rhythm. However, contrastive studies on German and English have previously led to the uncertainty whether these two languages are similar or different in terms of their prosody (see for an overview, e.g. Scuffil 1982). Grabe (1997: 157) explains that this disagreement might go back to different methodological considerations applied in different studies, e.g. the data were based on different analytical frameworks, there was no distinction made between phonetic and phonological levels of analysis, there were no directly comparable speech data, such as different texts and speaking styles. However, both languages rely on intonational cues when expressing contrastivity and signaling major

---

[26] Languages are not classified into the traditional rhythm classes of "stress-timed, syllable-timed and more-timed languages" (Pike 1945) anymore but instead are located somewhere along a rhythm continuum (cf. Gut et al. 2007: 12)

[27] Previous research suggests that in Czech the second syllable is frequently produced with higher f0 than the first stressed syllable (Palková and Volín 2003: 1784-1785; Volín 2008: 92).

information. In both languages, pitch is used in a systematic way to mark different syntactic structures (more on intonational phrasing in Section 2.6.5). As far as the f0 range is concerned, Mennen et al. (2007), for instance, found a wider f0 span but not f0 level for standard British female speakers compared to women speaking standard Northern German.

Just as all the other languages described here, Spanish is also an intonation language, albeit with a more syllable-timed rhythm. According to Bowen (1956: 31), Spanish has a similar tone inventory to English (i.e. high, mid, and low pitches). Similar to Czech, Spanish has syntactic resources to mark prominence, which may involve changes in word order rather than recourse to intonational cues. There is general disagreement on whether Spanish uses pitch accents to convey the meaning of contrastive focus, however evidence was found for a use of pitch accents to convey meaning in contrastive focus in Madrid Spanish (Face 2001). A general strategy used in Spanish as well as English is for focus to appear at the end of sentences, usually followed by a pause (Ortega-Llebaria and Colantoni 2014). The two languages also differ from each other in terms of pre-nuclear peak alignment, i.e. that Spanish peaks tend to appear on the post-tonic syllable and English peaks usually align with accented syllables (cf. Face 2001; Zárate-Sández 2016). Just like Czech and German, Spanish has generally been described to have a narrower f0 range than English (e.g. Kelm 1995).

However, research has not only uncovered differences between Czech, German, Spanish, and English, but has also determined regional differences within these languages, as well as in different varieties of English. Various cross-linguistic studies have attested to widespread regional variation in prosody (English: Grabe et al. 1998; Spanish: Sosa 1999; German: Gilles 2005; Ulbrich 2006a; Peters et al. 2015). Taken together, these similarities and dissimilarities constitute an interesting point of cross-linguistic comparison concerning L2 development, and a prediction of ease or difficulty L2 English learners may encounter when producing features in their L2. This will also allow for a test to Mennen's (2015) LILt working model previously presented in Section 1.4.3. This more general overview of differences and similarities between the prosodic systems of the four languages in this section will be complemented by more details in the respective sections on the prosodic features under investigation in the present study, i.e. tones and tunes in declaratives, intonational phrasing, f0 range, and uptalk tunes.

2.3 General problem areas of L2 prosody production and research gaps

Based on the prosodic similarities and differences specified above, Czech, German and Spanish learners of English are expected to show negative transfer in f0 range measures and positive transfer of the tone inventory. Additionally, due to the rather syllable-timed rhythm and reliance on word order instead of intonational cues in Czech and Spanish, negative transfer can be expected in intonational phrasing, i.e. longer phrases. In order to identify the major gaps in L2 prosody research, it is first necessary to provide an overview of what research has been done before and to what extent. Previous research on the prosody of Czech, German, and Spanish learner English has focused on a range of prosodic features such as intonation, f0 range, speech rhythm, foreign accent, accent placement, realization of tones, and intonational phrasing, among others. As becomes clear from the overview in Table 3, the analysis of prosodic features of CzE, GerE, and SpE speakers is a phenomenon that has been steadily gaining more academic interest since the 2000s. However, previous work has been limited to analyses on reading style and sometimes only focusing on question intonation (with the exception of Gut 2009 and Anufryk 2012, who provide large scale studies with a variety of L2 prosodic features). Table 3 provides an overview of studies on the L2 prosody production of CzE, GerE, and SpE speakers, excluding pure perception of foreign accent studies, as well as pure segmental and rhythmical studies, studies including only question intonation, and studies on child(-directed) speech (e.g. Riesco-Bernier and Romero-Trillo 2008; Riesco-Bernier 2012). If studies by the same authors were based on the same data set or a subset of the data, including the same linguistic and extra linguistic variables, they were not listed separately in the overview in Table 3, but they were listed under the most recent publication of the main author(s). Studies before 1975 were not included in the overview. The research presented here is only based on English non-native and native speakers and numbers on learners or NSs of other languages included in the studies are not provided in Table 3. In addition, the intonational models adopted in the studies are also provided.

Table 3. Overview of studies on L2 prosody from Czech, German, and Spanish learners of English. The number of non-native subjects in each study are highlighted in bold letters. Ordered from the most recent publications to older publications

Variety	Studies	n. subjects	Extralinguistic variables	Speaking style and text type	Intonational Model of Analysis (may be modified in the studies)	Tone Inventory and Meaning	Intonation/ f0 patterns	F0 range	Speech Rhythm	Foreign Accent (Perception)	Realization of nuclear pitch accents (stress)	Intonational Phrasing[28]	Accent placement
Mixed L1s (including either German, Czech, or Spanish learners)	Puga (2019)[29]	**10 GE-NNS, 10 SP-NNS**, 10 NS	Stay abroad	spontaneous speech (monologues and dialogues)	AM model ToBI		X					X	
	Kang, Rubin, Pickering (2010)	**6 SP-NNS**[30]	only male, comprehensibility and oral proficiency ratings	60 second response to summary task	Brazil (1997) discourse intonation	X	X	X			X		
	Gut (2009, 2017); see also for intonational phrasing only: Gut (2012)	**46 NNS**[31] 4 NS	AoL, LoR, instruction (INS), knowledge (KNO) and motivation (MOT)	read, retelling and free speech	AM model ToBI	X	X	X	X	X	X	X	X
	Pickering and Litzenberg (2011)	**11 NNS**		spontaneous speech - dialogues	Brazil (1997) discourse intonation	X	X	X					
	Pickering (2009)	**25 NNS** 4 NS		information-gap task	Brazil (1997) discourse intonation	X	X	X					

[28] There are studies with a focus on segmental features that is explained through prosodic structure/phrasing, which are, however, not listed in this overview (see Bissiri and Volín 2010 with Czech-accented English and Bissiri 2013 with German-accented English).

[29] The methodology I have adopted in Puga (2019) has changed in a few aspects in the present study. For instance, the tone inventory is different, i.e. “M” tone is not used anymore to indicate level. See Shobbrook and House (2003), Duběda (2014) and Anufryk (2012) for similar approaches of including a level tone.

[30] However, they did not specifically distinguish in their analysis between the different L1s. They also analyzed Chinese, Korean and Arabic speakers of English. Yet, they can be assumed to have a similar distribution as the subset of learners used in Gut (2017), including speakers of Germanic, Slavic, and Romance languages.

[31] It is not clear which L1 backgrounds the analyses in Gut (2009) were based on exactly. It is only stated that 17 different L1 backgrounds were identified in the group of 46 speakers of non-native English (cf. Gut 2009: 66). Gut (2017) uses the following learner groups: German (7), Hungarian (2), Spanish (1), Russian (1), and Polish (1).

Variety	Studies	n. subjects	Extralinguistic variables	Speaking style and text type	Intonational Model of Analysis (may be modified in the studies)	Tone Inventory and Meaning	Intonation/ f0 patterns	F0 range	Speech Rhythm	Foreign Accent (Perception)	Realization of nuclear pitch accents (stress)	Intonational Phrasing	Accent placement
L1 Czech	Skarnitzl and Rumlová (2019)	**10 NNS**	only female, strongly-accented	**read** (news)	N.A.			X	X				X
	Volín, Poesová, and Weingartová (2015), see for some first observations Volín and Skarnitzl (2010a, b)	**4 NNS** 20 NS	gender	**read** (news)	N.A.			X					
	Weingartová, Poesová and Volín (2014a)	**42 NNS** 8 NS	proficiency	Beginners and native speakers read (word list, short sentences and a limerick) Intermediate: **read** (news bulletin 500 words)	N.A.						X		
	Šimáčková, Kolářová, Podlipský, (2014)	**10 NNS** 1 NS		**read** (fiction story)	AM model ToBI							X	
L1 German (including Austrian and Swiss German)	Puga, Fuchs, Hudson, Setter, and Mok (2018)	**20 NNS** 10 NS		**read** (story)	Nuclear Tone Theory Halliday (1970)		X			X			X
	Mennen, Schaeffler, and Dickie (2014); see also Scharff-Rethfeldt, Miller and Mennen (2008), and Mennen (2007)	**21 NNS** 30 NS	only female, proficiency based on self-assessment, LoR and AoA	**read** (story)	AM model ToBI			X			X		
	De Leeuw, Mennen and Scobbie (2012)	**10 NNS** 10 NS	AoA, LoR	**read** (sentences)	AM model ToBI						X		X
	O'Brien and Gut (2010)	**10 NNS**	only female	read and semi-spontaneous speech	AM model ToBI						X		X
	Ulbrich (2012), See also Ulbrich (2010)	**30 NNS** 10 NS	exposure to L2 (yes 3 years + or no exposure at all), regional varieties of German, gender (male and female)	**read** (short text)	IViE[32] system (Grabe 2004)					X	X		

[32] Intonational Variation in English (Grabe et al. 2001).

L1 German (including Austrian and Swiss German)	**Studies**	**n. subjects**	**Extralinguistic variables**	**Speaking style and text type**	**Intonational Model of Analysis (may be modified in the studies)**	**Tone Inventory and Meaning**	**Intonation/ f0 patterns**	**F0 range**	**Speech Rhythm**	**Foreign Accent (Perception)**	**Realization of nuclear pitch accents (stress)**	**Intonational Phrasing**	**Accent placement**
	Ulbrich (2008)	**10NNS** 5 NS	Only female, exposure to L2	**read** (short text) and **(semi)-spontaneous speech**	IViE system (Grabe 2004)		X						
	Anufryk (2012); see also Anufryk (2008); Anufryk and Dogil (2009); Anufryk, Jilka and Dogil (2008)	**41 NNS** 12 NS	pronunciation aptitude (6 different groups from low, to highest aptitude), gender	read and spontaneous speech tasks in L1 and L2	AM model ToBI		X	X			X		
	Atterer and Ladd (2004)	**14 NNS** 6 NS	regional origin, L1 influence	**read**, free conversation	AM model ToBI						X		
	Gut (2000)	**6 NNS** 4 NS		read	AM model ToBI				X			X	X
	Grosser (1997), see also Grosser (1982, 1989, 1993)	**8 NNS** 2 NS	gender, L2 instruction (longitudinal)	read	AM model ToBI		X				X		X
L1 Spanish (including Latin American Spanish varieties)	Romero-Trillo (2014), see also Romero-Trillo and Lenn 2011; Romero-Trillo and Newell 2012; Romero-Trillo 2019 on subsets of the LINDSEI-SP and LLC[33]	**15NNS** **… NS[34]**	only female		Nuclear Tone Theory Halliday (1970)	X	X						
	Graham and Post (2018)	**10 NNS** 5 NS	proficiency (advanced C1 and Basic A2)	Naming task in a dialogue	AM model ToBI		X				X		X
	García Lecumberri, Cooke, Wester (2017)	**10 NNS** 10 NS	only female	spot-the-difference task - conversation	N.A.			X					
	Ward and Gallardo (2017, 2015)	**6 NNS** (+ 7 others) 20 NS (+ others)		spontaneous dyadic conversations	N.A.		X	X					

[33] London Lund Corpus (Svartvik 1990).
[34] Information on number of native speakers in the sample not discernable.

L1 Spanish (including Latin American Spanish varieties)	**Studies**	**n. subjects**	**Extralinguistic variables**	**Speaking style and text type**	**Intonational Model of Analysis (may be modified in the studies)**	**Tone Inventory and Meaning**	**Intonation/ f0 patterns**	**F0 range**	**Speech Rhythm**	**Foreign Accent (Perception)**	**Realization of nuclear pitch accents (stress)**	**Intonational Phrasing**	**Accent placement**
	Jiménez Vilches (2017)	**4 NNS**	only female, longitudinal	oral presentations	Brazil (1997) discourse intonation	X	X	X					
	Colantoni, Klassen, Patience, Radu, Tararova (2016b); see also Colantoni, Klassen, Patience, Radu, Tararova (2015)	**7 NNS** 7 NS	sentence-type	imitation task, complete the scenario task	AM model ToBI			X			X		
	Ortega-Llebaria and Colantoni (2014)	**14 NNS** 13 NS		answers in relation to a story	N.A.			X			X		
	Nava and Zubizarreta (2009)	**45 NNS** 30 NS	intermediate and advanced proficiency levels	read (story and scripted dialogue)	N.A.				X				X
	Zubizarreta and Nava (2011)	**47 NNS** 34 NS	intermediate and advanced proficiency levels	Scripted dialogue (and read story)	AM model ToBI				X				X
	Gutiérrez Díez (2008)	**15 NNS** 1 NS	4th year of ESO	read narrative text	Nuclear Tone Theory Halliday (1970)						X	X	X
	Gutiérrez Díez (2005)	**15 NNS** 3 NS	pre-intermediate learners (2nd year of Bachillerato)	read	Nuclear Tone Theory Halliday (1970)						X		X
	Ramírez Verdugo (2008)	**12 NNS** 12 NS		read speech and natural classroom discourse	Nuclear Tone Theory Halliday (1970)	X	X	X					X
	Ramírez Verdugo (2006b)	**10 NNS** 10 NS	19-22 year old Spanish speakers from Madrid, no time abroad, upper-intermediate proficiency, longitudinal study (3 years)	read (scripted short dialogues)	Nuclear Tone Theory Halliday (1970)	X						X	
	Ramírez Verdugo (2006a), see also Ramírez Verdugo (2002; 2003; 2005)	**20 NNS** 10 NS		read (scripted short dialogues)	Nuclear Tone Theory Halliday (1970)	X	X	X			X	X	X
	Ramírez Verdugo and Romero-Trillo (2005)	**10 NNS** 10 NS	only female	Read (scripted short dialogues) and spontaneous conversations	AM model ToBI		X	X				X	

L1 Spanish (including Latin American Spanish varieties)	**Studies**	**n. subjects**	**Extralinguistic variables**	**Speaking style and text type**	**Intonational Model of Analysis (may be modified in the studies)**	**Tone Inventory and Meaning**	**Intonation/ f0 patterns**	**F0 range**	**Speech Rhythm**	**Foreign Accent (Perception)**	**Realization of nuclear pitch accents (stress)**	**Intonational Phrasing**	**Accent placement**
	Romero-Trillo and Llinares-García (2004)	**2 NNS** 2 NS		read	Nuclear Tone Theory Halliday (1970)	X	X				X	X	
	Kelm (1995)	**?**	only female	scripted role play (dialogue)				X					
	Wennerstrom (1994)	**10 NNS**[35] 10 NS		read and free speech (picture description)	AM model ToBI	X		X					
	Archibald (1994); see also Archibald (1993a; 1993b)	**7 NNS**		read (sentences), word list	Idsardi (1992)						X		
	Backman (1979)[36]	**8 NNS** 4 NS	only male different sentence types	read (scripted & unscripted dialogue)	N.A.			X		X	X		X

[35] Heterogeneous group: four participants from Spain, five from Mexico and one from Argentina, with an intermediate proficiency in English.
[36] Even though Backman (1978) is often cited as one of the first empirical L2 intonation production studies, it suffers from methodological limitations, such as a lack of quantitative data and statistical analysis.

The studies listed in Table 3 and other studies across different L1s and L2 combinations (e.g. Willems 1982, with Dutch learners of English; Ullakonoja 2007, with Finnish learners of Russian; Busà and Urbani 2011, with Italian learners of English; Zimmerer et al. 2014, with German learners of French and French learners of German), have revealed many similarities concerning errors of L2 prosody, which could lead to the assumption that there may exist universal patterns in the acquisition process of L2 prosody. Errors in the production of L2 English intonation by speakers with different language backgrounds, which appear similarly across studies, are listed in the following (cf. extended list from Mennen 2007: 55):

- A narrower f0 range and smaller declination rate
- Problems with the correct placement of prominence
- Replacement of rises with falls and vice versa
- High proportion of falling tones
- Incorrect pitch on unstressed syllables
- Difference in final pitch rise
- Starting pitch too low
- Nucleus assignment to non-lexical items
- Problems with reset from low level to mid level after a boundary
- Intonational phrasing: tendency for short tone units (one to five words) and tendency for many tone units to be incomplete

Sections 2.5-2.8 address these prosodic deviances and findings from previous studies in more detail. Previous research on L2 prosody has given great insights into how prosody is produced and acquired, and what factors determine successful prosodic acquisition. After reviewing these studies, I will focus on the limitations of previous studies cited in Table 3 in order to highlight some general problems in this line of research. It is not to say that more controlled settings in which most experiments have taken place are not valuable, they are after all the basis of what is known about prosodic variation in L2 speech. Adding to this body of research, the present study extends the existing work on more controlled speech data by analyzing spontaneous speech. From Table 3 one can infer that studies on L2 prosody production of Czech, German, and Spanish speakers are still relatively scarce. With the exception of Gut (2009) and Anufryk (2012), each study on L2 prosody mainly focuses on read speech (stories, news bulletins, or word lists), and thus only includes one speaking style. Within these reading samples, usually only a few items (utterances) are selected (e.g. five sentences in Mennen et al. 2014). Due to the focus of these controlled studies on read speech, not much is known about L2 prosody in spontaneous speech (e.g. Gut 2009) and in interaction with NSs. Therefore, it becomes clear that the effects of entrainment by native speech have only marginally been investigated (e.g. including a model speaker in repetition tasks in read speech as in Zimmerer et al. 2015). By using read speech, interesting phenomena such as uptalk remain unexplored in L2 prosody research. Furthermore, intonational phrasing has not been fully investigated in read speech. Even though IPs are often used as the basic units that contain all f0 patterns, they are relatively fixed in read speech by punctuation. In spontaneous speech speakers are free to choose when to break up their utterances into IUs (depending on interviewer influence, if present). It is important to mention that readers and speakers have different pragmatic goals and different processing demands and many researchers have commented on the prosodic differences of read and spontaneous speech (Howell and Kadi-Hanifi 1991; Ayers 1994; Blaauw 1994; Sagisaka et al. 1997; Colantoni et al. 2016a, b; as summarized in Speer et al. 2003: 96).

Furthermore, the 38 major studies (59 total studies) based on different data sets listed in Table 3 have quite low numbers of participants (2-47 NNSs and 0-34 NSs, mean=16 NNSs (SD=13) vs. 9 NSs (SD=9), median=10 NNs (IQR=20) vs.7 NSs (IQR=10). They also mostly only look at one prosodic feature and one or two intonational dimensions (mostly distributional and/or realizational dimension), only compare non-native speech to one variety of English (mostly BrE and AmE, six out of the 59 studies use no NSs for comparisons), and include only very few (if at all) extralinguistic variables. Many studies, as shown in Table 3, only focus on one or two intonational features and might have, therefore, come to incomplete conclusions about the phonological system, since there is evidence that various prosodic as well as segmental features are acquired independently of each other in an L2, and the improvement in one feature cannot predict success in another (e.g. Wennerstrom 1998; Lleó and Vogel 2004 analyzed the relationship between segments and intonational phrasing in Spanish learners of German; Gut 2009 analyzed the relationship between consonantal processes, vowel reduction, speech rhythm and intonation in learners of English). 20 out of the 38 major studies (+ five mixed studies that also include SpE speakers, thus 25 in total) with different data sets are on SpE prosody with a clear preference of the Nuclear Tone Theory as their principal modal of intonational analysis. Ten out of 38 studies (+ two mixed studies that include GerE speakers, thus totaling 12 studies) investigated GerE prosodic productions with a tendency to prefer largely modified versions of the AM model and ToBI. For the CzE speakers, only four studies were conducted, with only one using the AM model of intonational phrasing and the others not mentioning a specific framework, because the analysis was based on phonetic analysis, e.g. f0 range or the prosodic realization of stressed syllables. The present study is unique in that sense in that it addresses all the research gaps mentioned so far (to the extent the data set used in the present study allows), by looking at the largest learner group so far (with the exception of Gut 2009), including three different learner groups from different L1 backgrounds. The influence of regional aspects has been ignored by most SLA studies, opting to treat AmE and BrE as a homogenous variety (cf. Ulbrich 2012: 4). For this reason, two different native varieties of English representing different dialects of English will be used instead of only considering BrE. While proficiency levels have been investigated in many previous studies (see Table 3), the only longitudinal studies for the L2 varieties investigated in the present studies were conducted by Grosser (1997) and Jiménez Vilches (2017). Proficiency levels and longitudinal aspects of language development cannot be investigated in the present study in a traditional sense, as will be explained in Section 3.3. Interestingly, the few longitudinal studies and semester abroad variable studies report that students do not improve their intonation (cf. Ramírez Verdugo 2003; Gut 2009). Gutiérrez Díez (2005, 2008) notes that different proficiency groups (pre-intermediate and more advanced students, respectively) produce strikingly similar results of intonation errors and do not see an improvement. However, as will be shown in Sections 2.5.3, 2.6.3, and 2.7.3, most learners of English (Czech, German, and Spanish) improve their prosodic patterns with increased proficiency, and that there might be a difference in the input the learners receive and a focus on prosodic training might be more important than mere exposure to an L2 (see Gut 2009).

How exactly learners differ in these prosodic aspects will be shown and explained in the following sections. All the similarities of errors found in the different studies summarized in the present section investigated different types of learners, with varying L1s and point to universal patterns in the acquisition of L2 prosody. However, following Mennen et al.'s (2010) approach, the present study does not view the acquisition of L2 prosody in terms of errors, but rather regards it in terms of its systematicity. Thus, the present study takes over a *learner variety* approach (Perdue 1993; Klein and Perdue 1997) in analyzing non-native prosody. Taken together, the present study is one of the first large-scale quantitative multifactorial analyses of non-native prosody to include various dependent linguistic variables and independent explanatory variables.

2.4 Prosodic features

The focus of the following sections (2.5-2.8) is on L2 intonation, i.e. tones and tunes, intonational phrasing, f0 range, and uptalk tunes in declarative utterances. First, prosody in English native speech will be treated briefly (definition, form, and function). Second, prosody in the learners' L1s is reviewed. Third, previous research is summarized on L2 prosody of Czech, German, and Spanish speakers in the respective prosodic features. Only where necessary, i.e. where no or little information on CzE, GerE, or SpE is available, will other studies including different L1s be discussed as well. Each section will conclude with findings on the form and function of the above-mentioned prosodic features in non-native speech and with information on how the present study plans to analyze them.

2.5 Tones and tunes in declaratives

The following section describes tones and tunes in NE speech. Before delving into this description, it is important to note that in the present study I adopt the following definition of tunes: "A phrase's tune or melody is defined by its particular sequence of pitch accent(s), phrase accent(s), and boundary tone" (Pierrehumbert and Hirschberg 1990: 277). Furthermore, when I speak of tones, I am referring to pitch accents and edge tones.

2.5.1 Tones and tunes in declaratives in native English speech

The present section takes only a brief look at NE tones and tunes, since they have already been described above in Section 2.1 within the AM framework. Section 2.2 also already provides details about their shape and distribution in comparison to Czech, German, and Spanish tones and tunes. Nevertheless, the description of tones and tunes in NE speech will act as the baseline of comparison for non-native tones and tunes in Section 2.5.3.

According to Pierrehumbert's (1980) model of AmE intonation, unmarked declarative statements or assertions are typically realized with a final falling tune, i.e. H* L-L%. Also, for BrE intonation, a falling tune seems to be the most frequent across different speaking styles (Nevalainen 1992). However, some speaking styles in AmE and BrE might exhibit different distributions of rising and falling tunes (Nevalainen 1992; Wennerstrom

1994: 415; Tyler 2019). Level tunes,[37] however, seem to be the most frequent tunes (approximately 50% of all tunes) in AmE across speaking styles (Wennerstrom 1994; Tyler 2019) and were found to be more frequent in BrE in interviews and depending on how distant a relationship is (Nevalainen 1992: 414). However, the more distant the relationships (not only socially distant but also in terms of physical distance in telephone conversations), the more rising tones are used (ibid.).

Besides stylistic variation, there is also considerable regional variation. Different varieties or dialects of English can vary as much from each other in their prosodic features as different languages, just as cross-speaker variation can also occur within dialects (cf. Grabe 2002; Grabe and Post 2002; Grabe et al. 2005; Arvaniti and Garding 2007; Clopper and Smiljanic 2011). For instance, in Belfast English and Urban Northern British (UNB) varieties, the default tune is predominantly rising in declaratives (Grabe 2002; Nance et al. 2015; Warren 2016: 25). Previous studies found not only phonological differences between the varieties of English, but also phonetic realizational differences, e.g. Southern California (SoCal) speakers using later f0 peak alignment than speakers from Minnesota (cf. Arvaniti and Garding 2007). These studies show that a three-way contrast[38] of high pitch accents, as proposed by Pierrehumbert (1980), is applicable to some AmE dialects and in some contexts (emphatic vs. non-emphatic contexts of use). Since the present study includes different NE dialects, this phonetic and phonological difference has to be accounted for in a detailed transcription system. Which tones were selected for the present study is discussed in Section 3.7.1.1, combined with the other ToBI labels for Czech, German, and Spanish.

Several studies have additionally determined that not only the regional heritage but also social identity can be encoded through prosodic patterns (e.g. Clopper and Smiljanic 2011; Ordin and Mennen 2017). However, gender/sex is rarely analyzed as a sociolinguistic variable. Those studies that included gender/sex as one of the explanatory factors found that women produced more and larger High Rising Terminals (HRTs) (Barry 2008), SAmE and Midland AmE women preferred late peaks (Clopper and Smiljanic 2011: 243), and SAmE females producing more H- phrase accents to connect successive ips than Midland females (Clopper and Smiljanic 2011: 243). However, female speakers of both dialect groups in Clopper and Smiljanic's (2011) study preferred the L-H% edge tone combination, which might suggest a gender difference in marking the relationship to successive IPs to mark continuation. Thus, a typical intermediate utterance in read speech from a SAmE female speaker would be L*+H H- and the typical intonation phrase: L*+H L-H%. Clopper and Smiljanic (2011: 244) suggest that this preference may reflect female speaker's attitudes and emotion toward the content of the utterance or could be interpreted as uncertainty. Thus, to summarize the information on the shape of tones and tunes in English declaratives, the default tune (H* L-L%) is used in unmarked declaratives; however, there are stylistic, regional, and social differences, as well as context-dependent features (e.g. Nevalainen 1992; Barry 2008; Ulbrich 2008; Clopper and Smiljanic 2011).

Tones and the manner in which they combine into tunes can fulfill many different functions. With tunes one can express attitudes and emotions, mark different speech acts or the beginning and end of turns in discourse, as well as focus on a particular part of the

[37] Wennerstrom (1994: 413-414) defines her plateau as a high phrase accent followed by a low boundary tone, which results in a flat pitch to signal continuation.
[38] See Ladd (1983) for an opposing view.

utterance. According to Warren (2016: 66), "[p]rosodic meaning has multiple layers and a single intonation contour can convey meanings simultaneously at more than one of these layers". It should be noted that consensus does not exist as to the number of functions of intonation. While Roach (2009) proposes four functions of intonation, Wells (2006) and Crystal (1995) identify six, and Lee (1958) suggests ten. Yet, the most widely accepted proposals concern a three-way division of intonational function, e.g. pitch, stress, and juncture (Trager and Smith 1951), indexical, linguistic, and discourse functions (House 2006), or tone, tonicity, and tonality (Halliday 1963). From her literature review of intonational discourse models, Pickering (2009) lists three major functions of intonation: information functions, discourse management functions, and relationship-building functions (cf. Pierrehumbert and Hirschberg 1990; Brazil 1997; Wichmann 2000; Wennerstrom 2001). Ladd (2008) only makes a two-way division, i.e. pitch (Halliday's tone) and relative prominence (Halliday's tonicity). According to Couper-Kuhlen and Selting (1996: 19), in naturally occurring data intonational patterns rarely carry a grammatical functional load in language use and each tone is assigned a pragmatic meaning within a given discourse context (cf. Kang et al. 2010: 556). In discourse, NE speakers tend to use low tones with a longer pause at IP boundaries; new topics are initiated with a high tone and mid tones are used for topic continuations (Nakajima and Allen 1993). Tyler's (2012) findings suggest that one of the functions of high tones is to mark coordination in discourse, similar to lexical items and syntactic structures being used to mark the relationship between sentences. According to Niebuhr and Ward (2018: 4), "intonational meanings are typically of a pragmatic nature and as such less tangible and definable than word meanings". Just to give an example of how different varieties of English may have different conventional or preferred intonational patterns for the same function, consider the forms of polite *yes-no*-question in BrE and AmE (example taken from Ladd 2008: 113-114). The utterance *Could I have the bill please?* would be realized with a falling-rising intonation pattern on *bill* and *please* in BrE, which comes across as condescending to American listeners. AmE speakers would use a high-rising tune instead ('*Could I have the 'bill /please*) to realize the same meaning. Similarly, Fletcher and Loakes (2006) found in the spontaneous conversational speech of AusE-speaking females that different tunes (i.e. mid-level plateaus, high rises, and fall-rises) are employed as floor-holders, and final falls tend to be used at turn-yielding positions. Generally, however, falls in declaratives can be said to signal finality and completeness; low falls can signal boredom or disinterest, while high falls might show interest and involvement (cf. Warren 2016: 10). Just as a simple fall, a rise-fall can also signal finality or completeness, while a rise-fall can also signal that an interlocutor is impressed (cf. Warren 2016: 10). According to Bolinger (1972: 28), the degree of conclusiveness signaled by a fall or a rise can be achieved by their extent, i.e. the deeper the fall, the more conclusive a statement is perceived to be and the steeper the rise, the more inconclusive it is perceived to be. Furthermore, the use of rises can serve as expressions of epistemic uncertainty (cf. Nilsenová 2006). According to the ToBI framework, the meaning of pitch patterns is compositional, i.e. each tone (accent) in a sequence contributes separately to the overall meaning and their pragmatic interpretation (cf. Pierrehumbert and Hirschberg 1990: 286-287; Cruttenden 1997: 64). A tune within an utterance is composed of a sequence of pitch accent(s), phrase accent(s), and a final boundary tone. While pitch accents convey information about focus, i.e. the highlighting of specific lexical items, phrase accents show how ips are to be interpreted in relation to each other, i.e. a high phrase accent (H-) signals cohesion between ips and a low phrase accent (L-) signals separation from subsequent phrases (cf. Cruttenden 1997: 110). Finally, boundary tones fulfill the same function as phrase accents but between IPs. Thus, a high boundary tone (H%) indicates topic continuation, while a low boundary tone (L%) indicates the completion of this part of the discourse (cf. Pierrehumbert and Hirschberg 1990: 287, 305). The present study sees the function of tones as compositional as in the AM approach and tunes as having a primarily pragmatic and social function.

2.5.2 Tones and tunes in declaratives in Czech, German, and Spanish

In line with the aforementioned small genetically and typological distance between German and English, many researchers have found only little or no difference between the tone inventories of German and English (Esser 1978; Fox 1981, 1984; Scuffil 1982; Grabe 1998). Both languages are intonation languages with a stress-timed rhythm and both use pitch in a systematic way to mark different syntactic structures. Czech and Spanish, on the other hand, have been described to be syllable-timed languages, which make use of word order instead of f0 to mark syntactic units (cf. Gutiérrez Díez 2012: 227). Generally, however, for English, Czech, and German it has been shown that pitch is used systematically for the marking of different syntactic structures (e.g. Petrone and Niebuhr 2014, with German L1 speakers; Pešková et al. 2018, with Czech L1 speakers) and prosodic information is used to disambiguate syntactically ambiguous utterances (e.g. Snedeker and Trueswell 2003; Snedeker and Yuan 2008; Jackson and O'Brien 2011). For English, there are a number of different descriptions of nuclear tones with different categories. Within the British School tradition (see Section 2.1) there are usually five[39] nuclear tones that are used for the description of English intonation, which include the following tones for both English and German: fall, rise, level (high, mid and low), fall-rise, and rise-fall. Other descriptions of BrE distinguish between seven main nuclear types, i.e. high-fall, low-fall, high-rise, low-rise, fall-rise, rise-fall, and level (Kingdon 1958; O'Connor and Arnold 1961; Cruttenden 1986, 1997). While there is disagreement whether German and English share the same nuclear tones or not when it comes to bitonal accents (cf. Trim 1964; Fox 1981, 1984; Raith 1986; as summarized in Gut 2009: 201), most researchers agree that the basic tone type distinction consists of the dichotomy of falling and rising tones (cf. Nevalainen 1992: 399).

The AM approach (see Section 2.1) and ToBI transcription system (see Section 3.4) have been applied to Czech, English, German, and Spanish. The following types of pitch accents, phrase accents, and boundary tones have been proposed for Mainstream American English (MAE; Beckman and Pierrehumbert 1986; Silverman et al. 1992; Beckman et al. 2005), Czech (CZ-TOBI; Pešková 2017), German (G-ToBI; Grice et al. 1996; Reyelt et al. 1996), and Castilian Spanish (SP-ToBI; Estebas-Vilaplana and Prieto 2008, 2010). Differences between the three systems are highlighted in bold letters in Table 4. There are two versions of German ToBI, i.e. one designed in Saarbrücken (Grice et al. 1996; Reyelt et al. 1996; Grice and Baumann 2002; Grice et al. 2005) and the other in Stuttgart (Mayer 1995). The present study chose to adopt the Saarbrücken version, because it follows the MAE-ToBI[40] conventions more closely than Mayer's (1995)[41] version. For SP-ToBI, the most

[39] Romero-Trillo (2001) extended Halliday's model of five nuclear tones by adding Tone 0, for cases (pragmatic markers) with no tone.
[40] Mainstream American English Tone and Break Indices (Beckman et al. 2005).

recent version was used (Estebas-Vilaplana and Prieto 2010). Table 4 provides a comparison of the various ToBI labels proposed and illustrates that MAE-TOBI and G-TOBI have the most ToBI categories in common. The dark gray cells show which tones all four ToBI systems have in common, while the light gray cells show the tones that at least two systems have in common.

Table 4. Overview of existing ToBI annotation systems in MAE, Czech, German, and Spanish

MAE-ToBI	**CZ-ToBI**	**G-ToBI**	**SP-ToBI**
Pitch accents	**Pitch accents**	**Pitch accents**	**Pitch accents**
H*	H*	H*	H*
L*	L*	L*	L*
L+H*	L+H*	L+H*	L+H* **and L+iH* and L+>H***
L*+H	L*+H	L*+H	L*+H
H+!H*		H+!H*	
!H*			
L+!H*			
L*+!H			
		H+L*	H+L*
Phrase accents	**Phrase accents**	**Phrase accents**	**Phrase accents**
L-		L-	L-
H-		H-	H-
!H-		!H-	
			HH-
			M-
			LH-
			LM-
			HL-
Boundary tones	**Boundary tones**	**Boundary tones**	**Boundary tones**
L%	L%	L-%	L%
H%	H%	H-%	H%
%H		%H	
		H-^H%	HH%
			M%
	(L)H%	L-H%	LH%
			LM%
			HL%
	!H%		
	!H:%		

Czech ToBI is based on a preliminary proposal by Andrea Pešková, who based her tone inventory on a south Moravian Czech dialect. So far, however, no standardized Czech ToBI exists (cf. Pešková 2017). See also Duběda (2011, 2014) for an earlier AM-based proposal of Czech tone inventories and nuclear contours. However, Duběda's (2014)[42] model was not considered in the present study because it seems too broad and rather too phonetic than

[41] For instance, Mayer (1995) includes M as a tone in the middle of a speaker's range, which is not allowed in MAE-ToBI. The M-tone is, however also part of SP-ToBI.

[42] Duběda's (2014) model includes a detailed annotation system of S- and M-tones, which do not appear in MAE-ToBI.

phonological in nature. Thus, Pešková's (2017) ToBI proposal was evaluated by me as more similar to the existing ToBI annotation systems of English, German, and Spanish intonation. The absence of phrase accents in Pešková's first proposal is explained by the analysis of quite short speech samples in her data set on which her proposed tone inventory is based. Based on her data, Pešková (2017) proposes that Czech is a phrase language (cf. Beňuš et al. 2014 for Slovak, which is very similar to Czech, and Hungarian), because it has fixed word-initial stress, and, according to Pešková (2017), is characterized by left-headed prosody. Phrase languages are languages with fixed lexical stress towards the edge of the word and can include accentual phrases (APs), which can be classified as a prosodic unit slightly larger than a prosodic word and smaller than ips within the framework of intonational phonology (cf. Jun and Fletcher 2014: 496). In a study by Beňuš et al. (2014) Slovak was found to make use of APs as a prosodic domain with a consistent f0 pattern (rise-fall) which was differentiated from the overall IP pattern. Even though CzE utterances may possibly be more similar to Czech APs, especially in more strongly-accented speech, for the sake of comparison to the other speakers groups, APs will neither be annotated nor analyzed and phrase accents will be included in the analysis of CzE instead.

As one can see from the overview of tones in Table 4, MAE-ToBI and G-ToBI propose the same phrase accents for both languages and are, on the whole, the most similar to each other, which was to be expected since both languages have a small typological distance. There is only one additional boundary tone for German, i.e. the upstepped high boundary tone (^H%). Nevertheless, the pitch accents are different from each other. While English has the single !H*, L+!H*, L*+!H, the German tone inventory additionally includes the H+L* pitch accent. As the overview shows, there are many different additional phrase accents and boundary tones proposed for Spanish. Despite the differences between the tone inventories of the four languages, there are many similarities in the way tunes are formed in utterances. As observable from the tonal configurations formed by the tone inventories laid out in Table 4, Czech declaratives consist of two main nuclear contours, i.e. (H)* L% and L*+H L% (the L*+H L% pattern being used only 15 out of 82 times by eight speakers from different dialects, i.e. Bohemian and Moravian: Pešková et al. 2018: 52). Similar results were found by Anufryk (2012) for native German (NG). Tonal configurations in Spanish and English declaratives are quite similar; however, they differ in pre-nuclear position, i.e. in that English has no late peaks (cf. Méndez Seijas 2018). Thus, just as in Czech, English declaratives are mostly produced in the following form: H* L-L%, while Spanish speakers would realize a late peak (L+H* L-L%). For all four languages, unmarked declaratives tend to be produced with a low boundary tone at the end of utterances (Sosa 1999: 140; Face 2005; Ulbrich 2006b: 370-371; Estebas-Vilaplana 2008; Anufryk 2012: 143; Colantoni et al. 2016a, b; Chamonikolasová 2017: 66; Pešková et al. 2018: 52).

Even though some similarities in tonal configurations of the four languages exist, differences were also found in the distribution of tones. When it comes to pitch accents, German speakers tend to prefer high pitch accents (H*) over low pitch accents (L*) (cf. Mayer 1995; Anufryk 2012: 120), the distribution of pitch accents also differs according to the speaking style used, i.e. with monotonal pitch targets (H* and L*) being more common in spontaneous speech (cf. Anufryk 2012: 126). Compared to BrE speakers, Mennen et al. (2012: 2258) found that German speakers preferred low pitch accents and this difference was directly noticeable at the outset of an utterance. When Czech is directly compared to English, the most frequent type of nucleus in scripted and non-scripted dialogues in both languages are falls followed by rises but falls being almost twice as frequent in English, while rises and falls were almost equally distributed in Czech (Chamonikolasová 2017: 66). For the non-terminal contours, the Czech and English speakers show the reverse patterns on rises and falls. While Czech speakers in non-scripted dialogues mostly produce rises (54.1%) and level contours (21.2%), along with some falls (15.6%), English speakers predominantly use falls (56.2%), fall-rises (20.7%), and some rises (13.2%) (ibid.: 67).

Even though the tone inventories of German and English seem to be particularly similar, both languages have differing phonetic realizations of several nuclear tones (e.g. truncation of IP-final accented syllables in German and compression in English: Grabe 1997: 160). Previous research has also provided evidence of cross-linguistic (German-English), as well as cross-varietal differences (Swiss German - Northern Standard German/ BrE dialects), which can be observed as L1 influence on L2 (German English) speech productions in the realization of nuclear pitch patterns (cf. see for alignment differences: Grabe et al. 2000; Atterer and Ladd 2004; Ulbrich 2012; for further intonational variation between German dialects see Petrone and Niebuhr 2014). Pešková et al. (2018) found that statements in Czech tend to begin with a higher f0 than in questions. Thus, Czech, which has been categorized as a phrase language, differs from languages like Spanish (Face 2005), German (Petrone and Niebuhr 2014, even though German questions can also have a final falling intonation), or English (see Warren and Daly 2005[43] with New Zealand English) (intonation languages), which make use of peak alignment and/or a high f0 peak to mark questions or statements (cf. Pešková et al. 2018: 53). Additionally, native Czech, relies on the prevalence of post-stress melodic rises (L*+H) (cf. Volín et al. 2015: 121). Compared to German speakers, AmE productions have been described to contain more tonal movements, more rises and falls, and more extreme endpoints of these tones (Jilka 2007: 90). The nucleus in English, Czech, and German is very frequently placed on the last lexical word at the end of IUs (Grabe 1998; Ulbrich 2004; Anufryk 2012: 143; Duběda 2014: 85; Chamonikolasová 2017: 53), whereas in Spanish the nucleus is placed on the very last word in an utterance regardless of its word class category (cf. Gutiérrez Díez 2005: 141). However, also nucleus placement elsewhere up to seven words away from the tone unit end for English non-scripted dialogic speech and the sixth word for Czech non-scripted dialogic speech were found in Chamonikolasová (2017: 53). However, these instances occurred only once in both corpora, and it was more common to find the nucleus in non-scripted dialogues on the last or second to last word (ibid.). Despite the differences in nucleus accent placement, there are also strong parallels in the phonetic realization of onset in English and Spanish, i.e. that it falls on the first syllable of the IP (cf. Gutiérrez Díez 2012: 224). Furthermore, in English and German, nucleus placement depends on givenness, i.e. given and accessible information is deaccented and produced with falls, while new information is accented and accompanied by rising tones (Brown 1983; Gut 2009: 250). However, Baumann and Grice (2006) and Pierrehumbert and Hirschberg (1990) claim that both given and new information can also be accented (i.e. receive a pitch accent) but different types of pitch accents are used for that purpose (i.e. low for given information and high for new information). The usage of givenness and accent placement is not a universal feature of all languages, while Germanic languages (e.g. German and English) seem to have

[43] Warren and Daly (2005) found that rises on questions started earlier than on statements.

similar or identical accentuation patterns, larger differences have been described for Romance languages (e.g. French, Italian, Spanish) from Germanic languages (Ladd 1996; Hirst and Di Cristo 1998; Hualde 2003). “A clear contrastive feature is that de-accenting old information is compulsory in English, whereas in Spanish it is neither compulsory nor very frequent” (Ortiz Lira 1994 as cited from Gutiérrez Díez 2012: 227). While in English sentence meaning is often modified by “tonicity”, Spanish relies more on changing word order (cf. Gutiérrez Díez 2012: 227). Thus, the present study will assume that while German and English speakers will exhibit similar patterns in signaling given and new information by the use of different tones, Czech and Spanish speakers will rather opt for word order. However, the present study did not distinguish between given and new elements.

Not only are tones used to signal the information status (given vs. new), but rises and falls can signal different pragmatic functions. Rises for instance can signal that the speaker wants to continue the turn and does not want to be interrupted and falls are used to signal utterance completion (cf. Möbius 1993: 19-20; Dombrowski and Niebuhr 2005, 2010). According to Grover et al. (1987), English uses large falls to signal continuation mid-sentence, while German and French use rises for that purpose and for termination tunes German uses steeper falls compared to English (see also Delattre 1963; Delattre et al. 1965). The latter finding, however, did not reach statistical significance in their study and was questioned by Chen (2003), who contends that BrE and German both use H% to signal continuation (see also Bolinger 1958; Pierrehumbert 1981: 991) but differ in the use of pre-boundary pitch accents. However, the results are moderately significant for the German preference over certain continuation patterns. Chen (2003: 1072) concludes that German listeners may not be “very keen on exploiting the phonetic shape of H% to signal different degrees of continuation in German” compared to BrE speakers. However, signaling these language-specific pitch patterns is not that simple, and they mostly only apply to carefully read isolated sentences, as in the early experiments by Grover et al. (1987).

To conclude, the differences in the tonal patterns found in previous studies might be due to methodological or stylistic differences rather than language-specific patterns. In addition, different tonal patterns can be found according to the context and formality of the situation. Spontaneous speech, in particular, does not always exhibit typical tonal patterns of continuation and termination. However, all four languages tend to produce low boundary tones in unmarked declaratives and high ones for continuation, with some systemic and distributional preferences for the reverse patterns of pitch and phrase accents. Therefore, some interference could be expected in the use of non-terminal declarative tunes, with SpE speakers possibly producing late peaks and more pitch accents in general, CzE and GerE speakers producing more low pitch accents, as well as more high pitch and phrase accents compared to BrE speakers.

2.5.3 Tones and tunes in declaratives in non-native speech

The knowledge about the tone inventory of Czech, German, and Spanish learners of English is quite limited. In fact, whether NNSs of English even have a different tone inventory than NSs of English is also still not entirely clear. Most studies that investigate non-native tone inventories propose the same labels for native and non-native speech (e.g. Gut 2009; Mennen et al. 2012), or settle for a compromise between two or more tone inventories based on different languages under investigation (e.g. Anufryk 2012). Gut (2009) found that all of the tones she proposed for English and German were also used in non-native speech. Therefore, the present study will assume a compromise between all ToBI labels (see Section 3.5 and 3.7.1.1). Even though Czech, German, Spanish, and English all have the same unmarked declarative form (H* L-L%) in statements, many researchers report on deviations of the respective learners (except Czech learners of English) in their distribution, combination, frequency, and realization of tones.

The type and distribution of tones in GerE and SpE speakers, very often deviate from the productions of NSs in similar contexts, and also between the interlanguages. Opposing results have been found by previous research on the distribution of tonal categories in free speech. While Gut (2009: 233) has found level tones to be most frequent (80%), and falls (10.24%), and rises (8.59%) occurring equally frequent, Wennerstrom (1994: 415) found SpE speakers to produce high and level boundary tones almost equally frequent (43% vs. 44%), while low boundary tones (13%) occurred mainly near the end of the text and before topic shifts. A substitution of rises with falls and vice versa has been reported for both GerE and SpE speakers (e.g. Backman 1979; Ramírez Verdugo 2002; Gut 2009; O'Brien and Gut 2010). O'Brien and Gut (2010) found that GerE speakers produce fewer pre-nuclear rises in their L2 than in their L1, i.e. the pre-nuclear accents are rather marked by falls. While a L+H* L-L% tune has been found to be the most common structure for both native and non-native English declarative IUs by Ramírez Verdugo (2006a: 17), the SpE speakers used a wider variety of pitch accent structures than the English natives: H*L (47.06%), L*+H L- (41.18%), and L*+H% (11.76%) (cf. ibid.: 22). Cross-varietal realizational differences were found when it comes to boundary tones. German learners seem to produce more high boundary tones, which could even be observed for some advanced learners (cf. Anufryk et al. 2008; Anufryk and Dogil 2009; Gut 2009). Ramírez Verdugo (2002), however, observed SpE speakers to produce more falls in statements.

A series of studies by Anufryk and colleagues showed that GerE speakers become more native-like in their tonal choice with increasing proficiency and that the difference is greater for female learners (cf. Anufryk 2008, 2009, 2012: 291; Anufryk et al. 2008; Anufryk and Dogil 2009; Graham and Post 2018: 9-10, with Puerto Rican Spanish speakers of English), while Ramírez (2003) did not find an improvement after a three-year stay abroad by their Spanish learners. Complex nuclei (i.e. bitonal tones or combinations of three tones) are very rare in native speech and even rarer in non-native speech (e.g. Grosser 1997; Gut 2009 (less than 1.5%); Mennen et al. 2010; Jiménez Vilches 2017). However, the use of tones (simple and bitonal tones) changes with increasing proficiency in the L2 to more native-like patterns (cf. Grosser 1997; Gut 2009; Mennen et al. 2010; Jiménez Vilches 2017). These studies found that falls are generally produced before rises, and simple nuclei are produced before complex ones. Learners were observed to use the pattern L* H H% most frequently for statements, and a decrease of the use of high boundary tones over time along with an increasing proficiency can be observed (e.g. Mennen et al. 2010: 321-322; Anufryk 2012). A study conducted by Ulbrich (2008) provides evidence that, even though highly advanced GerE speakers acquire regional intonation patterns after long-term exposure, they transfer these patterns into different speaking styles. By contrast, NSs of English are able to suppress regionally marked intonation patterns according to the formality of the speaking style. Ulbrich

(2008) concludes that intonation may not be acquired as a marker of stylistic variation until later stages of the SLA process.

Many studies report that NNSs of English not only show differences in the distribution of the type of pitch accents and edge tones, but also tend to produce more pitch accents than NSs (e.g. Archibald 1997, with Polish and Hungarian learners; Grosser 1997, with Austrian learners of English; Ramírez Verdugo 2002; Gutiérrez Díez 2005, with Spanish learners of English; Gut 2009: 245, with SpE speakers tending to produce more pitch accents than other learner groups). Not only an excessive number of nuclei has been reported, but also the placement on utterance elements that usually do not require an accent (e.g. conjunctions, prepositions, etc.) are found, which is neither common in L1 English, L1 German, nor L1 Spanish, and can therefore be labeled as developmental errors (e.g. Grosser 1997; Gutiérrez Díez 2005: 142, 2012: 224). Especially SpE speakers tend to accent the very last word in statements and in answers independent of its lexical category (cf. Ramírez Verdugo 2002: 130, 2006a; Romero-Trillo and Llinares-García 2004: 76; Gutiérrez Díez 2005; Nava and Zubizarreta 2009: 182; Zubizarreta and Nava 2011: 661). The wrong nucleus placement seems to be a typical pattern for learners in the early stages of intonational acquisition (Grosser 1997). However, some studies have also shown that the high frequency of pitch accents is also characteristic of intermediate and higher levels of L2 proficiency (e.g. Hiligsmann and Rasier 2002; Zubizarreta and Nava 2011: 661). In sum, the SpE speakers seem to be the most deviant learners in terms of overproduction of pitch accents, wrong accent placement, and accenting presupposed and given items (cf. Ramírez Verdugo 2002: 122), thus exhibiting both developmental (overproducing and wrong placement of pitch accents) and L1 transfer (accenting given items and wrong accent placement).

There is evidence for similar or equal phonological implementations of tones, but major differences have been found in the phonetic realizations of pitch accents (e.g. Ramírez Verdugo 2006a: 25; Anufryk 2012). Gut (2009: 248) also found the most evident difference between native and non-native intonation to be the phonetic realization of some tones. Romero-Trillo and Llinares-García (2004: 75) found that native and non-native speakers (L1 Spanish) make use of different mechanisms to mark prominence in unmarked utterances, in that NSs use amplitude and NNSs use duration. They suggest that this might be because their L1 (Spanish) is a syllable-timed language (ibid.). Interestingly, they found that the two groups make use of the opposite mechanisms when an item appears in marked position. For instance, fronted grammatical elements are marked by NSs by duration and NNSs use amplitude (ibid.). Additional differences in the phonetic realization were reported by Gut (2007: 10, 2009: 248), who found that the NNS's falls tend to be shorter (3.8 STs) than the rises they produce, i.e. the pitch movement stretches over fewer STs when falls are produced, which is in contrast to the NE productions (5.67 STs for falls), where the opposite pattern was observed. Similar deviances were found for Czech learners of English, in that they differed in the realization of stressed vowels, while L1 interference was also detected (cf. Weingartová et al. 2014a: 236). In BrE productions the stressed vowels were typically longer in their duration, exhibited a higher sound pressure level (SPL), higher f0, and a flatter spectral slope than on unstressed vowels. The performance of the Czech learners, however, seems to depend on their proficiency level, and the more advanced the learners are, the more native-like they realize the different parameters on stressed and unstressed vowels (cf. Weingartová et al. 2014a: 239). Thus, in Czech-accented English stressed syllables tend to be shorter and unstressed syllables longer than in native speech (cf. Weingartová et al. 2014a), which in turn can have an influence on intonational phrasing lengths in CzE.

Taken together, these studies suggest that a non-native interlanguage intonation system seems to exist for some learner groups. For instance, Ramírez Verdugo (2006b) found that NNSs approximate the NS usage of tones to fulfill various discoursal functions, while a quantitative analysis revealed differences with regard to the intonation system used, as has already been described above, e.g. NNS (Spanish in this case) do not distinguish between new and given information by the choice of tones they make and produce more falling tones (see also Puga 2019). Interestingly, in shorter utterances, Ramírez Verdugo (2006b: 524) found that tonality and tonicity patterns approximate the NSs' systems, and differences can be observed in longer utterances. Even though the general tune of the utterances seems to be similar, there are phonetic differences in low and high tunes within IPs.

In order to complement the existing studies on tones and tunes in non-native speech, the present study will adopt similar approaches as the studies presented above in the investigation of all four dimensions of intonation (distributional, systemic, functional, and realizational). Here, it is important to note that the present study includes the functional approach, which has been largely neglected in previous studies. Prosody is generally studied from a grammatical, pragmatic, and/or discoursal viewpoint. Hardly anything is known about the pragmatic functions of tunes in non-native speech, except that learners accent given information and deaccent new information (e.g. Wennerstrom 1994; Ramírez Verdugo 2002; Gut 2009) and that they use different tonal patterns and f0 range to signal different pragmatic functions of utterances in read speech (Ramírez Verdugo 2005, with Spanish learners of English who almost exclusively used falling tones to denote certainty and uncertainty in utterances). Therefore, the focus of the present study will be on the pragmatic (continuation, uncertainty, handing over turn, finality, etc.) and discoursal functions (speech acts: statements and answers, and topic shifts) that specific tunes fulfill. The present study does not focus on the grammatical function, where in the utterances accent are placed, which lexical categories are accented and which mechanisms are used to achieve prominence (amplitude, intensity, or duration). Duration features are only considered when taking a closer look at uptalk tunes (see Section 2.8). Furthermore, the syntactic completeness of a sentence was added on the annotation level, but it will not be further analyzed, as Ward and Gallardo (2017: 2) summarizes, "prosody is often determined more by the pragmatic function than the syntactic form, especially in dialogue". Therefore, the pragmatic function will take center-stage in the present study.

2.6 Intonational phrasing

The present section starts with a discussion of intonational phrasing in NE speech (Section 2.6.1), followed by NE speech as compared to Czech, German, and Spanish (Section 2.6.2), and then in non-native speech (Section 2.6.3). In the NE section, the structure of the paragraphs will be as follows: definition, external criteria and internal criteria for intonation phrase delimitation, form, and function. As will be shown in the present section, intonational phrasing in non-native speech is extremely under-researched. Likewise, studies on intonational phrasing in native speech are also relatively rare, since most researchers consider this part of intonation to be peripheral (cf. Ulbrich 2006b: 363). However, my view of intonational phrasing is that it is an important component of prosody that reflects a speaker's fluency and overall proficiency in the L2. Spontaneous speech, where speakers are generally free in the manner in which they break up their phrases, needs to be especially considered, in order to make the samples comparable. When I speak of intonation units (IUs), I refer to both levels of intonational phrasing, i.e. smaller intermediate phrases (ips), and larger intonation phrases (IPs). The term *intonation unit* (IU; Chafe 1988; Du Bois 1992; Selting 2005) will be the preferred term to refer to any form of prosodic chunking in this dissertation project. As will be explained in the following sub-sections, IUs are not only delimited by intonation, i.e. pitch, but by a variety of other prosodic phenomena. Other terms for this form of prosodic structuring have been proposed, e.g. *prosodic units* (cf. Szczepek Reed 2010), but the preferred term will be 'IU' because it is an established term with a commonly used abbreviation. Additionally, the terminology is consistent with the AM approach.

2.6.1 Intonational phrasing in native English speech

All models of intonation, as outlined in Section 2.1, propose the concept of a "basic intonational unit", which provides the "domain within which other intonational features occur" (Gut 2009: 207). Intonational phrases (IPs) can be defined "as segments of speech that occur within a single continuous intonation contour" (cf. Mennen et al. 2014: 327). Different researchers have used various names for IPs, e.g. tone unit, tone group, phonological phrase, breath group, or sense group, etc. (cf. Cruttenden 1997: 29-37). Since the present study follows the AM approach at describing intonational features, the term *intonation phrase* will be used to refer to such units, and a distinction is made between smaller ips and larger IPs. Similarly, in fluency studies, a *run*, which is defined as the amount of speech (typically measured in words, syllables, or phonemes) uttered between pauses (e.g. Lennon 1990: t-unit; Cucchiarini et al. 2000, 2002; Towell 2002; Freed et al. 2004), comes very close to the definition of an IP. According to Grice and Baumann (2007), "[t]he major difference between a 'run' and an intonation phrase is that in an intonational sense, speech is mainly divided into chunks delimited by means of intonation" (2007: 29). That IUs are not always delimited by intonation in natural conversations is discussed in Szczepek Reed (2010). Generally, IPs consist of one or more ips. Nevertheless, even if IPs and syntactic phrases are independent of each other, they often correspond (cf. Grice and Baumann 2007: 30). According to a corpus analysis conducted by Cruttenden (1997: 69), 40% of all IPs were coextensive with a clause. However, smaller grammatical structures were also reported (ibid). The syntax-prosody interface is discussed in Fernández (2010: 302), who writes that "prosodic phrasing is deeply driven by the syntactic content of an utterance" and "increasing syntactic complexity leads to different prosodic realizations". IPs in spontaneous speech do not necessarily have to be syntactically complete[44], but they should fulfill other important criteria, especially external and internal criteria that are outlined in the following.

[44] There are different approaches to segmenting IUs and how to account for syntax as well as other linguistic dimensions: While some suggest avoiding "syntactic thinking" during the segmentation of IUs (du Bois et al. 1992: 101), others incorporate the correlation between IUs with semantics and information structure (Chafe 1994), while some offer a more holistic approach to the segmentation of turn-constructural units, which include syntactic, prosodic, and action completion (Sacks et al. 1974; Schegloff 1996; Selting 2000, 2005).

Delimitating IU boundaries is a quite difficult task. There are certain external and internal criteria that can help to identify IU boundaries. Cruttenden (1997) suggests the following external phonetic criteria for the identification of IPs: pause, final syllable lengthening, anacrusis, change in pitch level and direction of unaccented syllables (f0 resetting). In her study on intonational phrasing in the LeaP corpus, Gut (2009: 225) divided each utterance of the 359 recordings (read speech, retellings, and free speech) of native and non-native speech into intonation- and interrupted phrases, with nucleus-presence and a pause being the major indicators of phrase boundaries. Yang (2004) found in her analysis of English speech that the proportion of phrase boundaries marked by pauses varies with speaking style (i.e. degree of spontaneity), i.e. only 35% of phrases in spontaneous speech from informal TV interviews are followed by a pause and 73.3% of phrases are followed by a pause in a narrative. Yang (2004) explains this phenomenon by the more structured nature of narratives (and more formal speech) and the use of pauses for different effects, such as emphasis, and informal interviews (more spontaneous, casual speech) being governed more by interactive cues and interruptions. According to Cruttenden (1997: 32), a pause, therefore, does not necessarily have to appear at every IP boundary, and the phonetic parameter pause is not a sufficient indicator for intonation boundaries and can only be considered together with other external or internal criteria. In a study by Clopper and Smiljanic (2011) IP boundaries were very often assumed when a pause longer than 100 ms appeared. However, in their data, at least for some cases, IP boundaries were marked by shorter pauses (shorter than 100 ms), or by no perceptive pause at all. In these cases, where no perceptible pause is present, a change in pitch level (also called f0 resetting) among unaccented syllables is responsible for such boundary perceptions and is an important indicator of IP boundaries (cf. Cruttenden 1997: 34; Grice and Baumann 2007: 29). In English one can observe a downward trend in f0 throughout utterances, called "declination" (Cohen and 't Hart 1967: 183-184). At the beginning of the following IP the pitch is picked up again, which is called "f0 resetting"[45] (Ladd 1988). Swerts (1997) observed that f0 resetting, pause length, and proportion of low boundary tones in spontaneous speech increase continuously with discourse boundary strength. Stirling et al. (2001) found that statement, question, and acknowledgement as discourse units were accompanied by small but significant upward f0 resetting and responses with a downward f0 resetting. Auran and Hirst (2004) found that f0 resetting depends on what type of IU follows, i.e. f0 resetting was found to be much higher at the beginning of IPs than for ips. Other external factors to delimitate IPs are anacrusis (fast delivery of unstressed syllables before the first pitch accent) (cf. Grice and Baumann 2007: 31) and syllable lengthening (Cruttenden 1997: 31-33). Despite these many phonetic correlates of the delimitation of IU boundaries, in many cases it remains difficult to decide whether an IU boundary is present or not. This is especially true for spontaneous and non-native speech, which contains hesitations, repetitions, and syntactically incomplete utterances. In cases where external criteria are missing, syntactic and semantic factors can be considered when making decisions about IU boundaries (cf. Cruttenden 1997: 36).

Furthermore, internal criteria of well-formed ips determine the delimitation of IU boundaries, i.e. each IU has to contain at least one stressed syllable with a particular pitch

[45] The resetting of F0 is analyzable in the f0 tracking in speech analysis software and is most visible after boundaries ending in a low tone. IPs ending in a high boundary tone do not show clear f0 resetting (cf. Gut 2009: 208).

movement to or from the accented syllable and only one phrase accent per ip and one boundary tone for IP (last ip) (cf. Cruttenden 1997: 35; Warren 2016: 11). Therefore, ips consisting of only unstressed syllables with no change in f0 level are not taken as separate phrases (ibid.). My criteria for IU delimitation for the present study also takes into account all the external criteria mentioned above, as well as the internal criterion of having at least one stressed syllable in one IU. More details on IU segmentation are described in Section 3.7.1.2.

Findings of Mukherjee's (2001: 67-78) intonation study include differences in speaking style, e.g. that monologues have twice as many ips as dialogues (7.71 vs. 3.35 ips per IP), and the higher the planning degree and the lower the level of formality, the higher the length of IPs (measured in ips). Clopper and Smiljanic (2011: 241) find no significant differences in the frequency of IUs in read speech of Southern and Midland AmE of male and female speakers. However, Clopper and Smiljanic (2011: 241-242) point to the assumption that intonational phrasing in read speech may be shaped similarly due to punctuation rules. Tench (2015: 31) contends that two-minute news bulletins contain 70-80 IUs, while more relaxed informal speech contains approximately 25 IUs per minute. On average, an English IP in native speech lasts for 1-2.5 seconds and is on average 1.9 seconds long (cf. Altenberg 1987: 22-23 as cited from Svartvik 1991: 560). It also contains about five words and very rarely over seven words (cf. Cruttenden 1997: 72). This, however, depends on the speaking style, e.g. read speech allows for longer IPs (in seconds and words) compared to conversational speech (cf. Cruttenden 1997: 72). In spontaneous speech, there is a mean of 4.8 words, including ips, and face-to-face conversations exhibit a mean of 4.3 words (cf. Nevalainen 1992: 419). As far as duration of utterances (measured in speaking rate in words per minute (wpm)) is concerned, conversational speech (university group discussions) has been found to be faster than public speaking in males, females, as well as different AmE dialects (Northern, Southern, and Western) (cf. Ray and Zahn 1990). Regarding the frequency and duration of IUs, the few studies that deal with intonational phrasing express different views on what an IU is and on the most effective research designs, which make comparisons across the studies quite difficult. However, what can be inferred from the studies is that there are speaking style differences and contextual factors that influence the frequency and duration of IUs. Nevertheless, based on previous research, gender does not seem to have an influence on intonational phrasing, compared to other prosodic features such as f0 range and tonal choice. However, keeping in mind that only few studies have been conducted on IUs, and the gender-variable is not always accounted for in previous research.

Not only stylistic variation has been observed in intonational phrasing and the duration of phrases, but also regional variation can be noted. Since intonational phrasing has yet to be systematically studied in AmE (with the exception of Clopper and Smiljanic 2011), studies on speaking and articulation rate will be briefly mentioned (the parallels between intonation and fluency have been described above), because fluency is seen as a related feature to intonational phrasing in the present study. According to Tottie (2002: 16), AmE tends to be spoken more slowly and loudly than BrE. Stereotypically, within the U.S.A., Southern American English (SAmE) is often described to be even slower than other American dialects (Preston 1998: 145; Baugh and Cable 2013: 365). Compared to New Zealand English, AmE (Connecticut) was found to be slower in terms of speaking and articulation rate, which is likely due to differences in vowel quality in these two varieties of English (Robb et al. 2004). Rob et al. (2004) also suggest that differences in rhythm may arise due to differences in vowel quality across the two varieties of English. Thus, the prosody of varieties of English may be due to vowel quality differences. Similarly, in accordance with the stereotype of slower speech in SAmE, Jacewicz et al. (2009: 246) and Jacewicz et al.'s (2010) found the articulation rate of speakers from Wisconsin to be faster than the one by speakers from North Carolina in read speech (by 8%), as well as interview speech (by 12.5%). Ray and Zahn (1990) could not confirm the stereotype of slower speech in three AmE dialects (Northern, Southern, and Western) of male and female speakers in two different speaking styles (conversational speech and public speaking), in that there were no significant differences in speaking rate due to region or gender. Clopper and Smiljanic (2011: 241) also disproved the stereotype of slower speech in SAmE in read speech by SAmE and Midland speakers, in that they observed no differences in the global speaking rate between the two dialects of AmE. Also no effects for pause duration have been found in their study; however, pause frequency seemed to be affected by dialect and gender. Southern male speakers produced more pauses per IP than the Southern female speakers or the male or female Midland speakers (ibid.: 241). However, it is important to note that Jacewicz et al.'s (2009) and Jacewicz et al.'s (2010) samples included more homogeneous groups of AmE speakers than Ray and Zahn's (1990) and Clopper and Smiljanic's (2011). In a later study Clopper and Smiljanic (2015) further analyzed segmental and global temporal characteristics in six regional dialects of AmE in the same reading passages, i.e. the articulation rate, pause frequency and duration, and vowel and consonant duration variability. They found that SAmE is characterized by a slow overall articulation rate, long pauses, and highly variable vowel durations. Compared to other dialects of AmE, the Midland and SAmE dialects were the slowest and the New England speakers the fastest in terms of speaking rate (Clopper and Smiljanic 2015: 5). Since SAmE is part of the analysis in the present study, these results are highly important, because if the Southern dialect is articulated slower due to segmental features and pausing, this might affect prosodic variation as well, especially intonational phrasing can be expected to be affected by this trend.

As is apparent from the discussion above, intonational phrasing fulfills a number of functions. Intonational phrasing is used to structure discourse into units of information in order to facilitate comprehension by the listener (Cruttenden 1997). Although the segmentation of IPs in spontaneous speech depends to a great extent on situational as well as individual factors, it is quite difficult to predict (compared to read speech), because speakers are free to choose when they break up an utterance. Nevertheless, "intonational phrasing is not arbitrary and it follows certain phonological rules" (Gut 2009: 210). How, intonational phrases are formed in the other three L1s (Czech, German, and Spanish), as well as in the respective L2 varieties, will be discussed in the next two sections.

2.6.2 Intonational phrasing in Czech, German, and Spanish

Intonational phrasing in all four languages seems to function in very similar ways according to different conventions in different speaking style, i.e. with IPs generally being longer in read speech than in spontaneous speech (e.g. Spanish/English: Gutiérrez Díez 2008: 329). Cross-linguistic comparisons of intonational phrasing in these languages are quite rare, since intonational phrasing is regarded as a rather peripheral phenomenon (e.g. cross-linguistic comparison of phrasing and accents in German and English: Batliner et al. 2001). Lleó and Vogel (2004: 80), however, distinguish between "demarcating languages" such as German and English, which tend to phonologically demarcate many IP boundaries. By contrast, Romance languages (Spanish, Italian, French) tend to phonologically group long strings into fewer larger constituents ("grouping languages" Lleó and Vogel 2004: 80). Their classification into these two groups is not meant to be an absolute classification, and both Spanish and German make use of feet, prosodic words, phonological phrases and intonational phrases, while intonation phrase boundaries are not obligatory but depend on the fluency and rhythm of speech (Lleó and Vogel 2004: 81-82). Nevertheless, in their study Lleó and Vogel (2004) only focus on feet and prosodic words in German and Spanish and did not look at the demarcation of phonological phrases and IPs.

According to Daneš (1957: 14), speakers of Czech have a tendency to segment their speech after pronouncing nine to eleven syllables. Palková (cf. 1994: 292-294) demonstrates that most utterances in Czech read speech as well as unprepared monologues consist of two or three rhythm groups and on average produce six to seven syllables in one utterance. Dankovičová (1998: 54) found IPs to be no longer than six phonological words in Czech. In direct comparison, Czech and English intonation have a lot in common, for instance IU length. Chamonikolasová (2017: 46) analyzed the IP length in Czech and English scripted and non-scripted dialogues. The IU length in Czech was 1-14 words long and 1-19 in English. However, in both languages and both speaking styles, the one-word and six/seven-word tone units were the most frequent in descending order. On average the IU length in words is 4.25 (2.57 SD) for Czech non-scripted dialogues, and 4.20 (2.84 SD) in English non-scripted dialogues (cf. Chamonikolasová 2017: 49), although the non-scripted dialogues are not directly comparable in a semantic sense. Another typical feature of Czech IPs appears to be phrase-final lengthening to mark prosodic boundaries (Dankovičová 1998; Volín and Skarnitzl 2007). In retrospect, Mennen et al. (2014: 316) analyzed IP length on the data set used in Mennen et al. (2012), and found that there are more IPs in German read speech (6.9 IPs) than in English read speech (5.7 IPs). The mean IP duration measured in seconds did not differ significantly between the two languages (1.38 s in English and 1.31 s in German). Mennen et al. (2014: 316) conclude that it was unlikely that the differences between the two languages can be traced back to large differences in IP length. Previous research by Nibert (2000, 2005, 2006) contends that Spanish intonational phrasing can be described just as English in AM terms, as IPs consisting of at least one or more ips (Sosa 1999 for opposing views on ips in Spanish; cf. Hualde 2002). According to Sosa (1999: 171-172) nucleus (or "intonation center") displacement is not possible in Spanish. Therefore, in order to signal focus intonationally in Spanish, utterances are divided into more tone units so that the focus is placed at a tone-unit boundary (ibid.). In corpus-based studies of contrastive intonational phrasing by Gutiérrez Díez (1982, 1983, 1995), which are cited in Gutiérrez Díez (2012: 220), the average IP length in spontaneous speech, i.e. televised and radio interviews, was 4.36 (English) and 3.94 (Spanish). Just as German and English, Spanish and English also show striking similarities in intonational phrasing. While the number of IPs in both languages does not seem to differ in spontaneous speech, Gutiérrez Díez (2012: 220) maintains that intonational phrasing in English has a much higher functional load than it has in Spanish. See also Esser (1978: 49), who also writes that the functional load of intonation is much higher in English than in German. While English makes use of intonational phrasing to disambiguate, Spanish prefers the subjunctive vs. indicative moods or changes in word order.

Taken together, these studies show that intonational phrasing in the four L1s is very similar, which leads to the assumption that it should not pose any interference problems to the learners, but rather should point to developmental problems, i.e. a lower proficiency level.

2.6.3 Intonational phrasing in non-native speech

Non-native intonational phrasing and information structuring are highly under-researched, as the overview in Table 3 Section 2.3 illustrated. While we know a few facts about intonational phrasing about a handful of Spanish, and even fewer German, learners of English, only one study has been published,[46] namely by Šimáčková et al. (2014), concerning intonational phrasing in Czech learner English, albeit with a focus on segmental and fluency features. The facts that we do know about intonational phrasing in non-native speech from previous research are summarized in the present section. One of the main reasons for the scarcity of research into L2 intonational phrasing is that non-native prosody is mainly studied in read speech, where due to punctuation, different interpretations of the text in terms of phrasing are quite unlikely. For instance, Tench (2015) contends that the vast majority of speakers (80%) agree on the division of the same text into IUs, although many counter examples have been observed for this in non-native speech (see Romero-Trillo and Llinares-García 2004; Gutiérrez Díez 2005; Gut 2007). Most of the studies cited in the following paragraphs very often do not specifically focus on intonational phrasing, but rather encounter it as a by-product, and, thus, mention it peripherally in their studies of other prosodic features. Gutiérrez Díez (2008: 341) found tonality errors to be the most frequent ones (5.8 on average), followed by onset errors (4.80), and tonicity errors (4.13), with quite low standard deviations for the last two (1.32 and 1.35), and a higher SD for tonality errors (2.95). This indicates that there is more variation in the production of IPs. However, it has to be stressed that all prosodic features are interconnected: “Many tonicity errors are prompted by concomitant tonality errors, which act as input to the former” (Gutiérrez Díez 2012: 227). In other words, intonational phrasing errors prompt the occurrence of pitch accent errors (cf. ibid). Gutiérrez Díez (2012) describes these intonational phrasing errors to be developmental and pitch accent errors as interference errors. According to him, pitch accent errors can only be fully accounted for if reference is made to the performance of IPs, since an excessive number of IUs leads to an excessive number of pitch accents (Gutiérrez Díez 2012: 228). This result, in turn, seems to be linked to the speaker’s lack of fluency (cf. Gutiérrez Díez 2005: 139).

At the fluency/accuracy interface, Gráf (2015: 82-83), for instance, demonstrates with a few examples from the Czech LINDSEI sub-corpus that something that would otherwise be regarded as an error can be corrected with the help of intonation:

(1) CZ005 - *“I wasn't sure. not at all”* - “whilst this appears in the transcription as a double negation error, the prosodic features make it clear that *“not at all”* is a separate non-erroneous clause here” (ibid.).

[46] However, research on intonational phrasing (prosodic phrasing) is underway in Czech news speech by professional and non-professional, as well as native and non-native, speakers of English and Czech, as recent conference presentations demonstrate (e.g. Volín and Galeone 2018 (Accents 2018 Lodz), 2017 (Accents 2017 Lodz)).

(2) CZ030 - *"so most that was another problem most of the time I stayed in the house"* - "checking the recording proves that this is an example of a false start which is resumed after the inserted clause" (ibid.).

(3) CZ033 - *"I don't think I have a .really .passion"* - "the pauses and intonation indicate that the speaker meant to use the disjunct *"really"* and not the adjective *"real"*. The idiosyncratic word order is not erroneous, the speaker interrupts her utterance, intensifies it, and then resumes it. This is a common feature of spoken language" (ibid.).

Examples (1)-(3) are characteristic features of spoken grammar, i.e. repetitions, reformulations, and incomplete utterances (cf. Biber et al. 2002; Miller and Weinert 2009; O'Grady 2010; Paterson 2011). These features of spoken grammar might influence the prosody of these utterances. For instance, examples (2) and (3) would influence prosody on the level of intonational phrasing, resulting in separating the utterance into two or more ips. This again would result in more pitch accents, since every ip requires at least one pitch accent. The restart of an utterance also very often involves a resetting in f0.

Overall, previous studies (mostly on read speech again) have found that NNSs of English tend to divide their IPs into a lot smaller ips than NSs (e.g. Gut 2000, 2007: 10, 2009; Ramírez Verdugo 2002: 122; Romero-Trillo and Llinares-García 2004; Gutiérrez Díez 2005, 2008: 341; Ramírez Verdugo 2006a: 21; Gutiérrez Díez 2012: 222; Puga 2019). Gutiérrez Díez (2005: 139) found the read speech produced by SpE speakers to contain five times as many tone units as in the NE productions. A similar result is reported in the study by Gutiérrez Díez (2008: 341) that 15 Spanish learners of English produced 70 IPs on average, as compared to 40 IPs produced by one BrE NS. It was also found that learners of English produce significantly shorter IPs (measured in words) in free conversations with NSs of English and in read speech (cf. Gut 2007; Mennen et al. 2010, with German learners of English; Šimáčková et al. 2014, with Czech learners of English; García Lecumberri et al. 2017: 182, with Spanish learners of English). Gut (2009: 228) found that the mean length of all IPs (in three speaking styles) is on average, 3.96 words in NNSs with different L1 backgrounds. However, it has to be mentioned that Gut (2009: 69) does not differentiate between ips and IPs, and the results of 3.96 words pertains to shorter ips. If one considers only the free speech part in Gut (2009: 229), IPs contained 3.36 words on average. In a corpus of read speech, Gutiérrez Díez (2008: 341) found that SpE speakers produce two words (SD: 0.37) per IP, while the BrE NS produced 3.30 words per IP. Gut's (2009 and 2012: 16) quantitative corpus analysis of intonational phrasing showed its high dependency on speaking style, e.g. learners were found to have a shorter mean length of run (MLR) measured in mean number of syllables per IP in story retellings than English NSs (5.4 vs. 11 syllables). In addition, NNSs of different L1 backgrounds produced many interrupted/incomplete phrases (15.81% of all phrases) (cf. Gut 2009: 228). Non-native speech is generally slower than native speech (e.g. Gut 2009; García Lecumberri et al. 2017: 182). Even though Spanish speakers use a great deal of elongations in their production when speaking in their L1, they increase the use of elongations when speaking in their L2, possibly due to L1 influence and/or combined with the "non-nativeness factor" (García Lecumberri 2017: 184). Thus, segmental as well as fluency features may determine IU frequency, which will be discussed in the next paragraphs.

Concerning the intonational-phrasing-fluency-link, Šimáčková et al. (2014) investigated whether speech connectedness increases with accelerated tempo in non-native pronunciation in Czech-accented English. They found that Czech learners of English organized their speech into fewer and longer IPs and reduced distinctiveness of word boundaries within IPs. Šimáčková et al. (2014) connect their findings to fluency measures (increased tempo) and, in doing so, they found evidence that tempo in read speech is reflected in prosodic phrasing, i.e. that the slowness of non-native speech leads to a greater number of prosodic breaks and more advanced and fluent speakers may run words together in fast speech and, thus, decrease the number of prosodic breaks. The AmE NS Šimáčková et al. (2014: 672) used in the different conditions of fluency, glottalization and intonational phrasing, found that the number of IPs dropped from 107 to 88 as a result of increased tempo in the self-paced reading mode. This was also the case in NG read speech, where slower speakers (self-selected/self-paced) generally produced shorter and more IPs and fast speakers produced fewer prosodic breaks (cf. Trouvain and Grice 1999). Fast native speech, on the other hand, has an effect on the prosody of an utterance and creates a specific particular pitch profile, such as an overall reduction of f0 range and a simplification of the tonal structure (cf. Fougeron and Jun 1998, with French L1 read speech). For a similar study with German read speech, only a simplification of bitonal tones to monotonal ones was observed, but not for an overall reduction of f0 range, and there is considerable interspeaker variability (cf. Trouvain and Grice 1999). By contrast, NSs of English have been shown to be capable of maintaining a constant alignment of f0 contours with segments even under changing speech rate conditions (cf. Ladd et al. 1999). Trying to increase one's fluency might come at the cost of more pronunciation, prosodic, or other errors (Derwing and Munro 1997; Skehan and Foster 1997: 207). Since two studies on fluency have been conducted with the exactly same learners as in the present study (Götz 2013; Gráf 2015), correlates of some these measures and prosodic deviances can be measured and compared. There is evidence that NNSs produce too many pitch accents (see Section 2.5.3). Results on non-native speech by Rasier and Hiligsmann (2007: 58) show that there is a correlation with fluency and prosody in that the more native-like the use of pauses in the utterance, the better the accentuation in non-native speech. Quantitative corpus analyses carried out by Götz (2013; LOCNEC and LINDSEI-GE), Gut (2009; LeaP), and Gráf (2015; LOCNEC and LINDSEI-CZ) showed that non-native fluency is characterized by a high pause ratio, i.e. a high proportion of pauses, both filled and unfilled, speech rate, and they produce shorter speech runs (MLR). In her study on native and non-native fluency, Gut (2009: 296) demonstrated a significant correlation with some intonational features. Götz (2013) and Gráf (2015) both found that the LINDSEI learners produced a lot shorter speech runs, with some variation in the different speaking styles. Both studies show strikingly similar results and report on 160 wpm for German non-native speech and 152 (Task 1: a monologue) and 157 (Task 2: an informal interview) wpm in Czech non-native speech. Thus, the learners in Götz's study were 40-50 words per minute slower than NSs of English in their speech rate in the same interview context. Gut (2012a: 16) also concludes in her study that learners of English produce many filled pauses, which are very likely to lead to slow speech rate and shorter IPs. Therefore, fluency is a crucial factor that influences the results on intonational phrasing in the present study.

At the segmental/supra-segmental intersection, glottalization is another delimitative feature of IPs, which can also give insights into L1 transfer, as the following studies suggest:

In Bissiri (2013: 236) GerE speakers glottalized 82% tokens of which the most frequent type of glottalization was creaky voice (CV) and these tokens were more frequent at phrase boundaries as compared to non-phrase boundaries. CV "refers to a vocal effect produced by a very slow vibration of only one end of the vocal folds; also known as vocal fry" (Crystal 2003b: 121). However, this phenomenon was reduced by 12% in speakers with a longer experience in an English-speaking environment (ibid.), but were still higher than English native productions in the same situation from Bissiri and Volín (2010). Rodgers (2000) found that CV is more frequent in spontaneous speech in German than in read speech (Trouvain and Grice 1999, with German read). The results from Bissiri (2013) and Bissiri and Volín (2010) on German and Czech accented-English might indicate an L1 influence of the frequency of glottalization, which ultimately influences intonational phrasing. As Bissiri et al. (2011: 165) summarize, Czech NSs use glottal stops regularly to mark prosodic boundaries, the Spanish use it quite infrequently and rather as an emphatic marker but also to mark prosodic breaks, and BrE speakers use glottal stops for both (boundary and emphatic marker), however, not as frequently as Czech speakers do. Yet, cutoffs and hesitations were excluded in Bissiri (2013: 236), since these are quite frequent in non-native speech and they can also cause glottalization, which is not necessarily a transfer from their L1. Therefore, CV instances will be added to the analysis, which may uncover possible L1 transfer phenomena related to intonational phrasing in the L2.

IP length has been shown to increase with experience in /exposure to the TL (cf. Punjabi and Italian learners of English: Mennen et al. 2010). In Mennen et al. (2010: 323) the length of IPs increased in each group after 30 months: from 2.3 to 2.9 words in the Italian learners and from 2.0 to 2.4 words in the Punjabi learners. Since Gutiérrez Díez (1982, 1983, 1995) found extremely similar IU lengths for English and Spanish in a CA, Gutiérrez Díez (2012: 223) concludes that most tonality errors (including the fragmentation of speech) could be labeled as developmental. Gutiérrez Díez (2012) expects these issues in phrasing to disappear naturally with increasing L2 proficiency, i.e. as learners gain prosodic fluency. Another type of developmental tonality error, which Gutiérrez Díez (2012: 223) notes, is that learners place tone-unit boundaries in the wrong syntactic junctions. For instance, English NSs will not expect a boundary between a main verb and its auxiliary. Gutiérrez Díez (2012: 223) calls this phenomenon "developmental" because he notes that Spanish children also make this mistake when learning their L1. Gutiérrez Díez (2012: 223) calls the fragmentation of speech and placing phrase boundaries at the wrong syntactic junctions a phonetic error, which are "a sign of distorted fluency and a feature of foreignness" and they generally do not cause a breakdown in communication as phonological errors would. As will be described in the following sections, NNSs typically tend to exhibit a narrow and compressed overall f0 range in comparison with NSs. The narrower f0 range limits the NNSs' ability to mark IP boundaries, especially when they start a new paratone/utterance (cf. Yule 1980; Swerts and Geluykens 1994; Pickering 2004: 31-32). Thus, it might be difficult for NSs to identify significant boundaries in non-native utterances and it might lead to confusion or misinterpretation of speaker intent. Interestingly, the ability of NNSs of English to delimit IPs from each other by means of their f0 range has been shown to be predictive of their rated oral proficiency in their TL (Wennerstrom 1998; Kang et al. 2010: 556). Pickering (2004: 31-32) and Kang et al. (2010: 561-562) found that NNSs were unable to use tonal patterns to create units of organization similar to those produced in native speech. Pitch peaks were often less pronounced in Chinese NNSs of English (Pickering 2004: 32). In my segmentation of IPs it was also difficult at times to hear if a new IP should be annotated, because sometimes there is no f0 reset and the f0 range does not change significantly for some NNSs. Therefore, the narrower f0 range might influence IP perception/annotation in return. This is one of the main reasons why f0 range is also analyzed. Pickering (2004) found intonational structure and phrasing to have a detrimental effect on the comprehensibility of lecture discourse. Most of the learners analyzed in the present study also expressed the intention to become teachers of English in the future, leading to the assumptions that their intonation might hinder comprehensibility and that it is, therefore, important to analyze its structure and its influence on the entire discourse.

As demonstrated above, not much is known about non-native intonational phrasing in Czech, German, and Spanish learners of English. In my opinion, IUs are indispensable units of analysis of other prosodic features, as outlined in the previous parts of this section. Therefore, I will analyze intonational phrasing in two different speaking styles from a pragmatic, topic structure, and discoursive perspective, by including also durational and fluency measures (i.e. IU length in words and seconds, and speech rate per IU).

2.7 F0 range

The present section is structured as follows: in Section 2.7.1, an f0 range definition, its shape, and functions are described. Section 2.7.2 directly compares English f0 range (form and function) to Czech, German, and Spanish. Finally, f0 range is discussed in view of non-native English speech in Section 2.7.3. However, it is necessary to first review the different measurements that have been proposed for the analysis of f0 range.

There are conceptional and methodological differences between approaches as to how f0 range is analyzed. Comparisons of f0 range are complicated by diverging methodologies, such as small participant numbers and a focus on read speech, but especially by the use of different measures to calculate f0 range.[47] Some of the most frequently used measures of f0 include the following:

- Long term distributional (LTD) measures based on the distribution of f0 within a given speech sample:
    - F0 level (mean, median, standard deviation), SD4, interquartile range, skew and kurtosis
    - F0 span: maximum minus minimum f0, 90th and 80th percentile
- Pitch dynamism quotient (PDQ)

Many studies using LTD measures have shown that English has a wider f0 range if compared to German (e.g. Mennen 2007; Mennen et al. 2007). However, as Mennen et al. (2012) maintain, there are position-sensitive differences between English and German f0 ranges rather than global differences (see also Wennerstrom 1994: 411). Thus, f0 range depends on where in the intonation phrase it is measured (see Patterson 2000). Many studies posit that

[47] See Patterson (2000), Mennen et al. (2012), and Mennen et al. (2014) for an overview of different f0 measures for cross-language and L1-L2 comparisons.

linguistic measures (i.e. measures of span and level are linked to specific landmarks in the f0 patterns) are more effective than LTD measures in capturing differences in f0 range across languages (e.g. Patterson 2000, Dutch/English; Mennen et al. 2012, German/English; Urbani 2012, Italian English/Italian/AmE). Patterson (2000: 124), for instance, found a linguistic measure excluding the first high boundary tones and accents from the analysis (M-L measures) to be more representative of f0 span, since the uncharacteristically high accents might skew the distribution.

Besides the different f0 measures mentioned above, different f0 scales have also been evaluated by Patterson (2000), i.e. Hertz (Hz), Equivalent Rectangular Bandwidth (ERB), and semitones (STs). In Patterson's (2000: 123) experiments of all scales of measuring f0 range (Hz, ST, and ERB), no differences were found in the correlation results, if level is represented in any of these scales. For span, however, Patterson (2000: 123) recommends using the musical STs, which have proven to best represent variation in span across speakers in his experiments. Nevertheless, many researchers use Hz for f0 level measures and STs for span measures (e.g. Mennen et al. 2007; Volín et al. 2015).

2.7.1 F0 range in native English speech

F0 range refers to the variation in f0 values that are used throughout utterances and are specified by speaker-specific minimum and maximum f0 limits. Patterson (2000:12) defines f0 range as the difference between a top-line and bottom-line from the f0 values, which is the physical correlate of pitch, used by a speaker during speech. For this reason, I refer to *f0 range* throughout the study, that is, what other researchers might have referred to as *pitch range*. The term *pitch range* is misleading in this context because it is not concerned with perceptual correlates of a particular f0 distribution (cf. Mennen et al. 2012: 2250). According to Ladd (1996, 2008), f0 range is to be considered a two-dimensional construct, with the dimensions of level (also called 'register'; overall f0 height) and span (the range or extent of f0 in a given speech sample). Level and span are two quasi-independent dimensions (see also 't Hart, Collier and Cohen 1990; Cruttenden 1997; Patterson 2000: 102; Ladd 2008). Regarding declarative utterances in English native speech and many other languages and dialects, usually a downward trend in f0 can be observed, also called "declination" (cf. Cohen and 't Hart 1967: 183-184). Declination means that the f0 is on average lower at the end of an IP than at the beginning.

Differences in f0 range between varieties of English, such as AmE and BrE, have been well documented. Some researchers propose that AmE displays less variety of intonation than BrE with a more "monotonous tone" (Baugh and Cable 2013: 365). Others have found that, for instance, SoCal English has a much wider f0 range than London English speakers in their productions of HRTs (Barry 2008). In addition, other dialects of AmE, such as SAmE, have been described as having a slower speaking rate (Preston 1998: 145; Baugh and Cable 2013: 365), but also more extreme f0 excursions have been observed (Feagin 1997: 134; Fox et al. 2013: 123, with North Carolina speakers vs. Ohio and Wisconsin). Feagin (1997: 134) found that white SAmE speakers use falsetto (the upper range of their f0), especially among teenagers, working-class and upper-class, male and female, for emphasis. However, some general tendencies hold true for all languages, i.e. females usually have higher and wider f0 ranges than males, and read speech is usually "livelier" (i.e. more f0 variation) than spontaneous speech (cf. Baken and Orlikoff 2000: 172: f0 is slightly higher in read speech), especially in short stories with different sentence types (but not true for more factual texts). F0 mean level in English male speech has been described to be between 106-128 Hz (Hirson et al. 1995; Hudson et al. 2007, with SSBrE speakers; Andreeva et al. 2014a, b; Volín et al. 2015, with Southeast England speakers), and 177-223 Hz for female speakers (Mennen et al. 2007, 2012; Andreeva et al. 2014a, b; Ulbrich and Mennen 2015, with one Belfast English speaker; Volín et al. 2015, with SSBrE speakers). F0 span in STs has been described to be between 9.2-15.1 STs for males (Andreeva et al. 2014a, b; Volín et al. 2015), and 6.4-13.9 STs in female speech (Mennen et al. 2007, 2012; Andreeva et al. 2014a, b; Volín et al. 2015). Most of the studies cited above, however, are based on read speech, with the exception of Hudson et al. (2007), who based their f0 values on simulated police interviews taken from the end of the interviews, which might explain their lower value of 106 Hz. Although, some studies report that task effects seem to have a great influence on f0 range, i.e. reading of a text/story versus a few sentences (cf. Mennen et al. 2007; Zimmerer et al. 2014; Colantoni et al. 2016a, b). Baken and Orlikoff (2000: 172), who review studies that have analyzed both read and spontaneous speech, summarize that no large differences were found between the two speaking styles and f0 variability. Furthermore, previous studies on English f0 range have shown that it exhibits a high interspeaker variability (e.g. Mennen et al. 2012: 2258; Andreeva et al. 2014a, b) and cross-dialectal variability as well (Shevchenko and Skopintseva 2004). For instance, Shevchenko and Skopintseva (2004) report on an extremely wide f0 range for Welsh and Scottish males, which was due to the high maximum values at the beginning of utterances. Equally high rising tunes have been described for Belfast English intonation (e.g. Ulbrich and Mennen 2015). Shevchenko and Skopintseva (2004) also report on differences in f0 range in different dialects of AmE, with AmE speakers from the South producing the widest f0 range in spontaneous speech (especially female speakers). There are also many different situational factors that might influence the shape of f0 range in NE speech, such as vocal fatigue, time of the day, ambient noise, the speaking style, the emotional content of the dialogue, the degree of familiarity between the interlocutors, etc. (cf. Braun 1995; Levitan 2014).

Many different functions have been ascribed to f0 range in native speech, i.e. to signal solidarity among speakers, discourse organization, agreement or disagreement (Ogden 2006), the attitude and the emotional and affective state of a speaker (cf. Ladd et al. 1985; Cruttenden 1997: 123; Kehrein 2002). However, one of the most important linguistic functions of f0 range, for the present study, is its organization of discourse units. F0 range and height are used to initiate new topics, mid f0 levels for topic continuations, and low f0 range for additional information or reformulations of previous utterances and at topic final boundaries (cf. Nakajima and Allen 1993; Swerts and Geluykens 1994; Cutler et al. 1997: 180-183; Wennerstrom 1998; Wichmann 2000: 130; Pickering 2004). This structuring of discourse is not accomplished by f0 range alone, but is a result of a combination of several prosodic features (i.e. f0, tempo, and pausing). For instance, prosodic fluency, i.e. the duration of pauses between utterances, can determine whether major (longer pauses) or minor (shorter pauses) topic shifts are undertaken (cf. Swerts and Geluykens 1994). In the present study, the use of f0 range will be analyzed further on the basis of discourse organization and how interaction is constructed in terms of turn-taking, floor holding, interruptions, and backchanneling[48] (cf. for an interactional approach to prosodic analysis: Couper-Kuhlen and Selting 1996).

2.7.2 F0 range in Czech, German, and Spanish

While some studies have investigated cross-language differences in Germanic and Slavic languages (e.g. Andreeva et al. 2014a, b), others compared concrete language pairs such as Czech vs. English (e.g. Volín et al. 2015), Spanish vs. English (e.g. Kelm 1995), and British English vs. German (e.g. Mennen et al. 2007, 2012). Andreeva et al. (2014a, b) have shown that languages (German, English, Polish and Bulgarian) can be categorized into different "pitch profiles" based on their f0 range. Differing f0 ranges have been described for NE and NG (e.g. Mennen et al. 2007), Spanish (e.g. Sánchez-Alvarado 2020), and Czech (e.g. Volín et al. 2015, 2017), and other languages, which led Andreeva et al. (2014a) to group languages into pitch profiles. In particular, they have shown that German and BrE speakers use lower f0 maxima, exhibit an overall narrower f0 span, and have a less variable f0 in comparison to speakers of Slavic languages, i.e. Bulgarian and Polish. However, gender showed to be an important predictor in f0 range, in that female speakers exhibited higher f0 values in all four languages (Andreeva et al. 2014a: 777). Early studies, in which English was compared to typologically different languages such as Spanish, Japanese, and Tagalog (cf. Hanley et al. 1966; Hanley and Snidecor 1967), found that the f0 range of AmE males exhibited the lowest median f0. However, more recent studies (Mennen et al. 2007; Andreeva et al. 2014a) showed that level and span measures are largely independent in the f0 ranges of the languages investigated, with there being speakers who have a wide span but differing levels, or speakers with very similar levels but differing spans.

Even though compared to Slavic languages, German and English seem to have more similar f0 profiles, when they are compared directly to each other, many differences can also be found. For instance, Jilka (2007: 85) showed that while the differences between peaks were very marginal and not statistically significant (average 216.5 Hz for Germans and 220.8 Hz for Americans), the baseline values were significantly lower for the AmE speakers (169.3 Hz vs. 182.4 Hz), implying that AmE speakers use wider f0 ranges (cf. Jilka 2007: 85). Kelm (1995) found that female NE speakers exhibit a wider f0 range when compared to female native Spanish speakers.[49] Estebas-Vilaplana (2014: 185) found lower f0 ranges for Spanish production compared to English ones. Interestingly, a recent study conducted by Sánchez-Alvarado (2020) found that Peninsular Spanish speakers have a wider f0 range in pitch accents realized in both informational and contrastive focus conditions compared to AmE speakers. Thus overall studies suggest that Spanish has a generally lower and narrower f0 range than English, but in certain conditions the opposite may be the case. Mennen et al. (2007) observed a 2.2 ST difference between German and BrE spoken texts measured by 80% range and an overall tendency of a wider f0 span but not level can be observed for standard

[48] The term backchanneling was coined by Yngve (1970). In this study I also adhere to backchanneling when I refer to small linguistic elements uttered by a listener that signal comprehension during speech. Romero-Trillo calls this feature "feedback" (2001: 536). A more detailed distinction between different types of backchannels was not undertaken (see Thompson (1996: 152) for a summary).

[49] Overall, Kelm (1995) found that native Spanish and English speakers do not differ significantly in their f0 range, however, the native English speakers were found to deviate more from that range than the Spanish speakers.

BrE female speakers compared to women speaking standard Northern German. In contrast to that, Mennen et al. (2012: 2255) found significant differences in both span and level f0 measures across the same groups of speakers. However, effect sizes were larger for f0 span and linguistic measures (compared to f0 level and LTD measures, respectively). Furthermore, contrary to anecdotal evidence of globally narrower f0 range for German speakers, Mennen et al. (2012) found that German speakers did not have an overall f0 narrowing over the course of an entire utterance, but instead exhibited a compressed f0 range at the beginning of IPs, which then expanded their f0 range in later parts of IPs.

Comparing Czech to English, Volín et al. (2015) also observed a 2 ST difference for 80% range in native Czech and English news reading of professional and non-professional speakers. When comparing the f0 levels, both male and female Czech speakers exhibited a significantly lower mean f0 than the NE counterparts (cf. Volín et al. 2015: 111). On the whole, Volín et al. (2015) found that native male as well as female English speakers not only produced significantly higher f0 levels, but also exhibited a wider f0 span and a steeper downtrend gradient than the Czech group. However, these results should be interpreted with caution, since only f0 range measures based on maximum and minimum f0 were used. F0 level and span for male speakers in native Czech (NE data in the same data set is put in brackets) has been described to be 107 Hz (120 Hz) and 6.1 STs (8.1 STs) measured in the 80th percentile range (Volín et al. 2015) in read speech, and 117.3 Hz (Skarnitzl and Vaňková 2017) in spontaneous speech. This is compared to 165 Hz (190 Hz) and 5.2 STs (7.1 STs) measured in the 80th percentile range in read speech spoken by female speakers of Czech (Volín et al. 2015). These results are extremely similar to the findings on NG f0 level, i.e. 115-121 Hz in German (Künzel 1989; Jessen et al. 2005; Andreeva et al. 2014a, b; Zimmerer et al. 2015). The results on German female speakers, however, differed considerably in a higher mean f0 value (pitch level) of 206-218 Hz (Künzel 1989; Mennen et al. 2007, 2012; Andreeva et al. 2014a, b; Ulbrich and Mennen 2015; Zimmerer et al. 2015). This result for female English speakers reported on in Section 2.7.1 is highly compatible. Scharff-Rethfeldt et al. (2008) found in read speech and retellings that female German monolingual speakers (f0 median in retelling: 194.6 Hz) exhibit a significantly lower f0 level than their BrE counterparts (f0 median: 316.7 Hz). For f0 span for female speakers of German Mennen et al. (2007) reported a value of 6.2 STs, which is within the frame of data previously reported on. With different measures for f0 span, however, Zimmerer et al. (2015), Andreeva et al. (2014a, b), confirmed the tendency for female speakers to always receive a slightly lower or equal value. Following these values, it seems that Czech male speakers receive the lowest f0 values and English female speakers receive the highest values of all four languages. Despite the conflicting results and different measures used in some studies, one can, nevertheless, assume that female speakers in all four languages under investigation in the present study will produce a higher f0 level most of the time.

Besides social (gender) differences, as mentioned above, many other aspects, such as regional and stylistic variation, have to be considered when analyzing prosodic variation. Regional variation has been shown to have an influence on f0 ranges in different regional accents of English (e.g. Grabe et al. 2000m in four different varieties of BrE; Shevchenko and Skopintseva 2004m with different BrE and AmE speakers). Grabe et al. (2000), for instance, showed that languages do share phonological implementations of f0, but differ in their phonetic realization and vice versa. Similarly, regional prosodic differences in German dialects (e.g. Ulbrich 2005, 2006a), Czech (see for a discussion: Uličný and Prošek 2013; Duběda 2014), and Spanish (e.g. Sosa 1999) have also been found. Nevertheless, speaking style has been analyzed more widely and different f0 ranges have been described cross-linguistically. Generally, a narrower f0 range is reported in spontaneous and free speech than in read speech (Wennerstrom 1994: 408, with native and non-native English speakers; Gut 2009: 249, with NE speakers; Skarnitzl and Vaňková 2017: 11, with Czech speakers). However, no such a distinction in f0 range between speaking styles (read, retelling, and free speech) has been found for NG (cf. Gut 2009: 249). In line with this, Baken and Orlikoff (2000: 172) point to small differences in their analysis of f0 range in read versus unscripted speech. Interestingly, the differences reported for German and BrE in Mennen et al. (2007) have been found to be most apparent in longer sentences and the effect disappears in sentences of a short duration. Hirson et al. (1995) showed that, in the speech of male speakers of English, f0 is consistently higher in read speech than in free speech. Hollien et al. (1997) found similar results with read speech exhibiting higher f0 means than in spontaneous speech; however, there was considerable interspeaker variability. For the present study, this means that not only does one have to pay attention to social, regional, and stylistic variation, the measures used for f0 range are just as important as the length of the utterances analyzed and position at which the f0 levels are extracted and measured.

2.7.3 F0 range in non-native speech

Previous studies have shown that regional variation in the native as well as the TL (e.g. Ulbrich 2012: 5) can influence L2 prosodic production. This further complicates the analysis of non-native f0 range. However, as identified in the previous sections on f0 range, there are many general linguistic tendencies that can be found cross-linguistically. As in all previous sections, the studies that are immediately relevant to the present study will be explained in more detail, while studies on other L2s (i.e. besides CzE, GerE, and SpE) will only be mentioned peripherally. The present section zones in on the question of whether learners of English reach target approximations or deviate from NSs in respect to f0 range. This also entails examining the possible reasons, i.e. L1 influence, proficiency levels, differing f0 measurements and study designs, along with extralinguistic variables. It will also be determined what is already known about the functions f0 range fulfills in non-native speech, and whether it generally approximates to the target norm.

Studies on L2 f0 range of Czech, German, and Spanish learners of English are quite rare, which is why it comes as no surprise that only very little is known about approximations in and deviations from the NS target norm. F0 range in non-native speech is, on the whole, narrower than native speech, irrespective of the L1 (e.g. Backman 1979; Wennerstrom 1994, 1998; Kelm 1995: 443; Ramírez Verdugo 2002, 2006a; Gut 2009; Urbani 2012; Volín et al. 2015; Orrico et al. 2016; García Lecumberri et al. 2017; Skarnitzl and Rumlová 2019). Many studies attribute their results pertaining to a narrower f0 range to uncertainty,[50] lack of

[50] At the same time, Volín et al. (2017) contend, “[w]e should be cautious about the term *uncertainty* used above, though. Most probably it is a complex affective and cognitive structure that will be difficult to define and measure. (This might be further complicated by the fact that the subjects often unconsciously deny the phenomenon’s existence, or the opposite - they consciously claim it while in reality they do not possess it.)” (Volín et al. 2017: 62).

confidence, or even moderate anxiety (cf. Mennen 1998: 18, with Dutch learners of Modern Greek; Zimmerer et al. 2014: 1037, with French learners of German and German learners of French; Volín et al. 2015: 121, 2017: 62, with Czech learners of English and English learners of Czech), rarely are any other explanations for this finding offered (e.g. for notable exceptions see Wennerstrom 1994; Gut 2009). One might even suggest that structural differences between the prosodic organizations of an L1 and L2 may explain the narrow f0 range; however, studies of f0 range in bilinguals, which show different patterns by gender, point out that structural differences might not always be the right interpretation (cf. Ohara 1999; Ordin and Mennen 2017). Nevertheless, Kelm (1995: 443) found that Spanish and English speakers speaking in their L1 do not significantly differ from each other in terms of their f0 range and that both groups reduced their f0 range when they spoke non-natively. In contrast to the narrow f0 range findings, García Lecumberri (2017: 189) found only a slightly smaller f0 range in SpE conversations, with no observed difference in f0 maxima; however, natively spoken speech exhibited a somewhat lower minimum. Mennen (2007) found that female GerE speakers in her sample (n=12), who live in the UK, neither differed in their span nor level across the two languages in read speech (reading in German and English). However, she observed that there was a tendency for a higher level in the learners' English, which, however, did not reach statistical significance (p=0.059) (ibid.: 68).

Especially with increasing proficiency in the L2, learners seem to develop more native-like phonetic implementations of f0 (e.g. Wennerstrom 1998, with Chinese learners of English; Orrico et al. 2016, with Italian learners of English). For instance, Anufryk (2009, 2012: 287; Anufryk and Dogil 2009), with GerE speakers, encountered a higher degree of phonetic variability as well as higher f0 within higher-aptitude speakers, in particular, female speakers, which approximated the NE target productions. Gut (2017: 217) found that a higher phonological competence level (albeit, only measured by foreign accent ratings) before a stay abroad or a training course predicted a larger gain in f0 range in her story retelling data. Jiménez Vilches (2017) reported on Spanish learners having increased their mean f0 range after a nine-week oral discourse intonation training. Other studies based on various L1 backgrounds have found that reduced f0 range does not always occur in the L2, especially with increased L2 proficiency (e.g. Mennen 2007, with German learners of English; Aoyama and Guion 2007, with Japanese learners of English who had a larger f0 range for some target sentences and content words; increased L2 proficiency after stay abroad: Ullakonoja 2007; Mennen et al. 2014; Zimmerer et al. 2014). In other words, the higher the proficiency, the more target-like the productions in f0 will be. However, no clear correlation with student proficiency and f0 range measures could be found in other studies (e.g. Hincks 2004, with Swedish learners of English; Zimmerer 2015, with German learners of French and French learners of German). In an experimental set up, Anufryk (2012) found that the L2 speech of less proficient/talented speakers exhibited a smaller degree of variability in their L2 English, compared to subjects with highest and high pronunciation aptitude, especially female speakers. However, the intonational variability by average-proficiency/aptitude speakers (especially by females) was close to those of the high proficiency groups in several conditions. These results were confirmed only for the phonetic level (PaInTE parameters) of the analysis and not for the phonological level (ToBI annotations). Orrico et al. (2016) observed a narrower f0 range for their Italian learners of English. While f0 level and nucleus accent placement improved with increasing proficiency, the f0 contour (mostly level contour) and the realization of nuclear accents (higher variability between high and low accents, while NSs are quite consistent in preferring H as a nuclear accent) seems to be particularly susceptible to fossilization. In Mennen et al. (2014: 323) the measure H*i-H* showed a significant correlation with self-assessment of the proficiency level in the L2, which shows that speakers who assessed themselves to have a more native-like performance were also more similar to the NE target productions of H*I and H* pitch accents. However, Mennen et al. (2014) suspect for the other two measures and the overall good performance of most of the 21 NNSs, that the learners in their sample might be so successful because of their more advanced level of proficiency and experience with the L2, since they live in the UK (Mennen et al. 2014: 324). However, the variables AoA and LoR did not show any significant correlations with any of the selected f0 range measures in Mennen et al.'s study (cf. Mennen et al. 2014: 323). In sum, previous studies on f0 range in L2 speech and the extralinguistic factors of proficiency/AoA and LoR show different and varying results, i.e. either more native-like productions with increasing proficiency or experience with the L2, partial improvement, or no change at all with increasing proficiency.

Besides narrower and no larger differences in the f0 range, *overhitting* (Gut 2009: 239; Anufryk 2012) of the L2 target as well as incidents of *underhitting* (cf. Volín et al. 2015) have been found. Orrico et al. (2016: 119) call these phenomena *prosodic drift*, with speakers moving away from their L1 but in an aimless way. For instance, Anufryk (2012), found that some female speakers of high pronunciation aptitude realized a higher mean f0 level in their English compared to German productions and productions of female NSs of English. Volín et al. (2015)[51] found that f0 measures of span were even lower in CzE than those measured in native Czech, but the f0 measures of level of CzE lies in between the native Czech and English reference groups. In Volín et al. (2015), while f0 level in accented-speech was found to take intermediate values between the two L1s (Czech/English), the f0 span was the narrowest. Both studies conclude that processes other than L1 transfer are taking place. Taken together, previous studies give relatively consistent evidence for a narrower f0 range in L2 productions and show a target approximation with increasing L2 proficiency. However, often f0 range results are based on very small participant numbers and do not take into account that f0 range exhibits considerable interspeaker variability (cf. Mennen et al. 2014: 308). Gut (2009: 239), for instance, reports that out of 46 NNSs analyzed in her data set,[52] 18 exhibited a maximum f0 range within the range of the NSs' in the retellings and 12 in the readings. Five learners even exhibited a wider f0 range than the average NSs' in reading passage style (cf. Gut 2009: 239). Mennen (2007) also observed interspeaker variability, i.e. more than half of the German speakers in her study showed a higher level in their English than in their German, and only two speakers exhibited a wider span in their English (Mennen 2007: 68-69). Even though some of the findings by Mennen (2007) and Volín et al. (2015) might not have reached statistical significance, they, nevertheless, suggest that f0 *level* measures are generally closer to the TL norm, and f0 *span* measures deviate more strongly from that norm for most learners.

Most previous studies rely on global f0 measurements of maximum and minimum values for level and span. However, position-sensitive measures within IPs have shown to be

[51] However, Volín et al.'s (2015) findings are based on a very small sample of four CzE speakers only.
[52] As Mennen et al. (2014: 307) note the L2 learners in Gut's study had widely differing L1s with varying degrees of distances from the TL, which make it difficult to interpret the reduced f0 range as L1-influence.

more revealing. Mennen et al. (2014) found through linguistic measures of f0 range that their female L2 speakers (n=21) often produced language-appropriate f0 range values and often approximated to the target, despite some observable deviations from the target. Both target deviations and approximations were found to be position-sensitive (Mennen et al. 2014: 323). For the measure of initial high pitch accent to the final low boundary tone, i.e. H*i-FL (comparable to the global LTD measure of f0 range), they found that learners deviated significantly from NG but not the NE values (cf. Mennen et al. 2014: 323). For the f0 range measure H*-[53] FL (every high non-initial pitch accent to the final low boundary tone), they found that the learners' values did not significantly differ either from NG or from NE values. Therefore, the learners were somewhere in between the values for NG and NE (cf. Mennen et al. 2014: 323). Finally, their third measure of the learners' values for H*i-H* (the initial high pitch accent to any of the next high pitch accents) differed significantly from both NG and NE values (intermediate values between the two languages) and was the only f0 range measure which significantly deviated from the performance of the NE speakers (Mennen et al. 2014: 324). Thus, they explained the result for H*i-H* by a possible L1 influence (ibid.). In sum, linguistic measures of the f0 range are position-sensitive and can give insights into L1 transfer, as well as provide a more accurate picture than global f0 range measures.

When it comes to L1 influence, studies of highly proficient bilinguals have shed more light on stylistic variation, as well as prosodic convergence. Interestingly, monolingual German and English speakers tend to change their f0 range according to different text types. Highly proficient bilinguals of German/English, on the other hand, did not make such a clear differentiation between two different text types when speaking English, and thus an L1 influence (German) could be observed (cf. Scharf-Rethfeldt et al. 2008: 126-127). However, the bilinguals did not resemble monolingual German nor monolingual English standards in f0 range realization in both text types (reading and retelling), thus taking an intermediate position between the two (ibid.: 127). These findings suggest that bilinguals as well as highly proficient learners of English may have changed their L1 f0 range in the process of acquiring an L2, and this has implications on the comparisons to be made on the f0 range productions of L2 learners and those produced by monolingual speakers.

Furthermore, prosodic entrainment may occur during specific phases in discourse and changes over the course of a conversation (e.g. De Looze et al. 2014, with Japanese speakers; Zellers and Schweitzer 2017: 2336, with German speakers). A study conducted by Ordin and Mennen (2017) with English/Welsh male and female bilinguals suggests that there is gender-specific variation in the use of the f0 range rather than a cross-linguistic difference. Their findings show that female bilinguals use different f0 ranges in their two languages, while male bilinguals do not make such a distinction (Ordin and Mennen 2017). In order to explain their findings, they speculate that "males need to manifest their individuality" and females are more likely to "conform to the society norms" (cf. Ordin and Mennen 2017: 1503). Previous research suggests that women tend to accommodate more and agree more often with the interlocutors in order to foster positive social relationships and to reduce social distance and enhance cooperative communication (cf. Namy et al. 2002; psychological studies: Eagly 1978; Cacioppo and Petty 1980; as summarized in Ordin and Mennen 2017:

[53] In Mennen et al. (2014) H*i and FL only occur once in every IP but H* can occur many different times and each one of these occurrences was included per IP.

1503). Not only do women accommodate more than men in their speech to other speakers, but they seem to be prone to do so when their "interlocutors" are male (cf. Namy et al. 2002: in their study they are not interlocutors but "shadowees" in a shadowing task). Ordin and Mennen (2017: 18) conclude that, for a bilingual, switching one's f0 range is a learned "behavioral pattern", which has been historically viewed as constituting typical "gender expressions". Similar results have been obtained by previous cross-linguistic research on the f0 range of Japanese/English bilinguals. Only the female speakers considerably increased their f0 level when speaking Japanese compared to their speech in English, while male bilinguals exhibited little prosodic variation in f0 levels across the two languages (Ohara 1999; Loveday 1981). F0 in English by male and female speakers is less distinguishable than Japanese speech of both sexes (Loveday 1981). Ohara's (1999: 113) study also found that the formality of the situation (communicating with a professor vs. a friend) has an effect on the f0 (higher pitched voice) but only for female bilinguals. However, in English, a wider f0 range has been reported to express politeness by both male and female participants (Loveday 1981). Perception-based research on phonetic convergence during imitation in single-word shadowing task suggests that AmE male speakers entrain more than females when they are in a subservient role of conversation, i.e. the instruction receiver in a map task (Pardo 2006: 2388). This was confirmed by Thomason et al.'s (2013: 753) study on prosodic entrainment in dialogues with students and pre-recorded and synthesized tutor voices. They found that male speakers entrained their loudness minimum and maximum features significantly more but not for f0. Based on her literature review of conflicting results, Levitan (2014: 86), argues against a straightforward relationship between gender and entrainment and suggests that the social context of an interaction (e.g. familiarity, topic, social status, emotional content, etc.) is more likely to have an effect on entrainment than gender. Especially the entrainment of f0 seems to depend on the interaction context, while other prosodic features such as intensity and speaking rate seem to be more automatically entrained (Levitan 2014: 109). The studies cited in this paragraph have shown that f0 range entrainment in bilingual and non-native speech may depend on the gender (of the speaker and the interlocutor), speaking style, formality, and the role a speaker takes in a conversation and that these differences can give insights into L1 transfer in f0 range. However, female speakers are generally more prone to entrain to their interlocutors in either their L1 or L2.

As can be seen from the previous studies of non-native f0 range, there is still much more research to be conducted in order to get a more comprehensive picture of the L2 f0 range. Most previous studies did not control for gender or included only male or female participants in their studies (e.g. Backman 1979; Ulbrich 2008; Mennen et al. 2014). They also had very small participant numbers (e.g. Volín et al. 2015). In addition, other speaking styles besides read speech were rarely investigated. In order to compensate for interspeaker variability, a larger group of speakers will be compared in the present study. The present study aims at characterizing the f0 range of Czech, German, and Spanish learners of English in comparison to the AmE and BrE NS groups. These results will be compared to the findings of L1 Czech, German, and Spanish speech[54] in order to attest for a possible L1 influences. The learners of English in this study are expected to have an overall higher f0 level (or no

[54] This is based on the literature review and not my own study since native recordings are not available for the data.

change at all) and a narrower f0 span in spontaneous speech compared to native BrE and AmE speakers. Based on the literature review, it can be expected that the f0 range differences between native and non-native groups will be position-sensitive. Therefore, both LTD and linguistic measures are employed to analyze the differences in the f0 range across the groups, as well as to see whether these differences manifest themselves globally or in certain positions within the f0 contour. Another aspect to consider is IU length, since no significant difference was found with increased proficiency and f0 range for shorter utterances (Zimmerer at al. 2015). Ramírez Verdugo (2002: 130) also found that NNSs approach the native target in terms of tonality and tonicity patterns in short utterances and there are greater differences in longer utterances. Thus, more variation in the f0 range is expected in longer and more coherent texts, i.e. discourse (cf. Zimmerer et al. 2015). Backman (1979) also found that the f0 range was wider for SpE speakers in shorter declaratives (18Hz wider) than their AmE counterparts, but 11 Hz narrower in longer declarative utterances. A CA of the Spanish recordings of these same subjects did not explain the subjects' problems in their f0 range of declaratives (Backman 1979: 260). The utterances in my data of the various speakers, are of varying lengths, and could also vary in the f0 range. It will be tested whether IP mean duration differs significantly from each other for the three learner groups. If it does, then appropriate utterances will have to be selected for comparison. In this study f0 range measures of the interviewers directly before an interviewee IP will be compared and convergence will be investigated. Thereby, I hypothesize that, during interaction with an interviewer, the interviewee will converge his/her f0 range for certain tonal targets to that of the interviewer, or possibly vice versa. The strongest effect is expected on the f0 range, as opposed to the other prosodic features (e.g. IP structure and length) and therefore, only entrainment effects on f0 range are investigated. In the present study, I will also show that speakers' local f0 values are influenced by the preceding f0 values of the interviewers.

2.8 Uptalk

The open/closed dichotomy of high (associated with open meanings, i.e. asking a question or signaling continuation) and low tones (associated with closed meanings, i.e. signaling completion or handing over turns) proposed by Cruttenden (1997: 163) has been challenged by various uptalk researchers (e.g. Horvath 1985; Nilsenová 2006; Ritchart and Arvaniti 2014; Tyler 2015, 2019; Armstrong et al. 2015; Arvaniti and Atkins 2016; Levon 2016, 2018), who proposed that the dichotomy needs to be broadened to take into account another interactional dimension, i.e. the use of uptalk. In previous research, uptalk has been described as a tool to help structure conversational interaction (e.g. Levon 2016, 2018). Previous research (e.g. Ritchart and Arvaniti 2014; Tyler 2019) has shown that there is a systematic variation in the way uptalk rises are used and that "it is problematic to discuss rises as an undifferentiated category" (Tyler 2019: 17). Therefore, in the following section, uptalk will be first defined, followed by an examination of its shape and a discussion of the functions it fulfills in NE speech and in comparison to Czech, German, and Spanish (Section 2.8.2). In Section 2.8.3, it will be shown what is known about the occurrence and use of uptalk in non-native English speech productions.

2.8.1 Uptalk in native English speech

High tunes in questions and low tunes in unmarked declaratives are the canonical form in most languages and varieties (see Sections 2.5.2 and 2.5.3 for the canonical tunes on declaratives). However, in some varieties of English, such as Belfast English, the default nuclear pitch accent is predominantly rising in declaratives (cf. Grabe 2002; Warren 2016: 25). Uptalk, which can be defined as rising tunes used with declarative utterances, has been observed in different varieties of English around the globe, i.e. in Australian (e.g. Fletcher et al. 2005), New Zealand (e.g. Britain 1992; Daly and Warren 2001; Warren 2005), American (especially SoCal English: e.g. Ritchart and Arvaniti 2014; Armstrong et al. 2015), and British English (e.g. Shobbrook and House 2003; Barry 2008; Armstrong et al. 2015; Arvaniti and Atkins 2016; Levon 2016, 2018).[55] Levon (2018), however, argues that HRTs in London English are a relatively recent arrival in the variety of English (the first mention of uptalk in BrE is made by Cruttenden (1994: 130, 1986)). However, Cruttenden (1997: 133), also states that many northern cities in Great Britain use rising tones more extensively within BrE. Uptalk in AmE is usually a phenomenon predominantly associated with the West Coast of the U.S.; yet, its use is common all over the U.S., especially among younger speakers. Thus, uptalk usage has been attested not only for SoCal English but also in the South (e.g. Ching 1982; McLemore 1991; Clopper and Smiljanic 2011), Midland (Clopper and Smiljanic 2011), Midwest (e.g. Prechtel and Clopper 2016), and North (Armstrong et al. 2015) of the U.S.. List 1 shows a comparison of structural and functional features of uptalk in AmE and BrE. Generally, more uptalk instances were found in selected dialects of AmE than BrE ones, where sometimes none at all were found (Barry 2008: 18). Uptalk has been attested globally among the young and particularly among young female speakers of English (Bryant 1980; Horvath 1985; Guy et al. 1986; Courtney 1996; as summarized in Warren 2016: 96). While some studies have found that younger speakers use more uptalk than adult speakers (e.g. Guy et al. 1986; Fletcher et al. 2005), other studies suggest that uptalk is not exclusively used by female (e.g. Levon 2018 has even shown that male white speakers of Southern British English (SBrE) use uptalk more often than white women; Armstrong et al. 2015), nor by younger speakers (e.g. Arvaniti and Atkins 2016). The present study adopts Warren's (2016: 2) working definition of uptalk: "a marked rising intonation pattern found at the ends of intonation units realized on declarative utterances, and which serves primarily to check comprehension or to seek feedback".

The studies of different varieties of English have shown that there are different tunes used for uptalk. Uptalk typically takes the following form on declaratives: it begins on the last tonic syllable of the intonational phrase and the intonation rises sharply through to the end of the phrase (cf. Guy et al. 1986: 25; Ritchart and Arvaniti 2014: 7). The description of the speech segments in which uptalk occurs also varies, but the focus is mostly on uptalk instances in IP-final position (Armstrong et al. 2015), although some also report on uptalk turn-medially at the end of ips (Barry 2008; Arvaniti and Atkins 2016). Most uptalk tokens in Arvaniti and Atkins (2016) were, however, utterance-final and 3% of all uptalk instances were ip-final. Among others, the following uptalk realizations and (pragmatic) functions have been reported across varieties of English:

---

[55] See Wilhelm (2016) for a multidialectal study of English.

List 1. Uptalking tunes in different varieties of English

- Australian English: L* H-H% (statements) or H* H-H% (mostly questions) (Fletcher et al. 2005: 396)
- American English:
  - SoCal English: L* L-H% (statements) and H* H-H% and L* H-H% (questions), L* H-H% and L* L-H% (also used for floor-holding and confirmation requests) (Ritchart and Arvaniti 2014)
  - SoCal: L* H-H% and H* H-H% (Barry 2008: 111, 140)
  - Midland: L* H-H%, L* L-H% and H* L-H% (Prechtel and Clopper 2016)
  - Northern: L* L-H%, L* H-H% and H* L-H% (Prechtel and Clopper 2016)
- British English:
  - Urban North British (including Belfast, Liverpool, Tyneside and Birmingham): "rise-plateau-slump" (Cruttenden 1997: 133-134) represented in AM terms L*+H H-L% or L+H* H-L% by Ladd (2008: 127)
  - SBrE: H* L-H% (floor holds, marked dip) and H* H-H% (confirmation requests (indirect questions to negotiate common ground with the addressee)) (Arvaniti and Atkins 2016)
  - London English: L* H-H% and H* H-H% (Barry 2008: 111, 140)
- Different varieties of English: L* H-H% or L* L-H%(38.85% of the total number of HRTs) L+H* H-H% (36.31%; conversational narratives, Wilhelm 2016: 3)

The overview in List 1 will serve as a comparison for the uptalk tunes in the present study. As can be seen, I adhere to Ladd's (1996, 2008) suggestion to include low onset (L* H-H%) tunes when analyzing uptalk, instead of only adopting Halliday's (1967) original definition of HRTs, which includes only high nuclear tones and high rising boundary tones (H* H-H%) that are at least 40% higher than the nuclear accent. As demonstrated in List 1, the different varieties of English can have different phonological categories, although a L* H-H% (low-how tune) is commonly found in most varieties. Differences in the phonetic realization (f0 range, alignment with syllables, and duration) of these different varieties of English have also been reported. For instance, the phonetic shape of uptalk has been described to be more variable in London English than in SoCal English (Barry 2008: 182). Generally, it has been assumed that intonational variation is not necessarily due to differences in phonological inventories but rather phonetic realizations of these phonological categories (cf. Fletcher et al. 2005: 401-402). As summarized by Levon (2018: 1), previous studies on uptalk (including his studies) focused on the following phonetic properties: rise excursion size, rise dynamism, and the alignment of the rise onset with the nuclear syllable. Guy et al. (1986: 25) found that rises in AusE typically showed an f0 increase of at least 40%, and the rises began on the last tonic syllable of the IP and continued sharply through subsequent syllables. Britain and Newman (1992: 4) found that HRTs in New Zealand English range from slight (less than 50 Hz) to dramatic (more than 100 Hz). Even for dialects of AmE (Massachusetts and SoCal English) with similar inventories and distribution of uptalk tunes, phonetic variation in uptalk has been observed in the duration and f0 range, i.e. SoCal female speakers produced the longest rises and females from both regions in the US produced steeper slopes than their male counterparts in narratives (cf. Armstrong et al. 2015). The typical declarative uptalk rises in London

English are described as moderately large and dynamic with a fairly late onset (cf. Levon 2018: 17). In sum, there is considerable regional variation in phonological inventories as well as phonetic realizations for uptalk tunes in NE.

When it comes to the frequency of uptalk, the proportion of uptalk reported in various studies on AusE varies from 69% to 72% of final rising tunes in AusE female speech; 11% to 13% were uptalk examples of the Tarrington and Melbourne speakers respectively (cf. Fletcher and Loakes 2006: 46). Similarly, McGregor (2006) found that uptalk occurred 11% of times in her study of discourse study (as cited from Fletcher and Loakes 2006). Fletcher (2005: 1382) found an occurrence of about 26% (195 instances out of 766 IPs) for final rising tunes in AusE from a map task. In Levon (2016), out of 7,351 declarative IPs by male and female Londoners (n=42 speakers, including only the white speakers) elicited from spontaneous information interactions from friendship groups, 719 (9.8%) contained uptalk instances. However, individual speakers in Levon (2016: 142) reached percentages of 3.5-41.2% uptalk usage. In Levon (2016) 18.9% of men and 10.5% of women use uptalk in their speech, which is closer to the percentages previously found for AusE. Interestingly, in mixed-sex talk, men used uptalk 18.9% of the time but in single-sex contexts men only used it 6.2% of the time, while no significant effect on mixed- or single-sex groups was found for female speakers (Levon 2016: 142). Levon's (2016; 2018) research has shown that the frequency of the occurrence of uptalk does not only depend on categories such as gender, but also information status (new and given), pragmatic function, and text type. Levon (2018) also shows nicely that variation in form (phonetic realization) does not necessarily map onto variation in function (contrary to previous research),[56] at least in the case of SBrE. In contrast to the findings on SBrE uptalk, all other studies cited above found systematic correlations between different uptalk shapes and the distinct functions that they are used to perform. This finding will be explained in more detail in the next paragraph.

In addition to differences in form, uptalk across varieties of English is used for different purposes. On a more negative note, uptalk is usually associated with uncertainty and lack of authority (cf. Warren 2016: 115). As described above, uptalk is a marked linguistic feature, which besides providing/sharing (new) information (cf. Tench 2003: 217), fulfills several other functions, such as an emotional and attitudinal function (James et al. 1989: 15, as cited from Warren 2016: 47), the reduction of social distance (Bradford 1997: 34-35), check on listener comprehension (Guy et al. 1986), interactional functions, seeking a response, eliciting hearer acknowledgement (Wichmann and Caspers 2001: 4), and signaling information structure (Warren 2016: 47-65). In addition to that, Shokeir (2008: 17) lists a variety of pragmatic functions that have been identified for uptalk in previous research, i.e. continuation, uncertainty, deference, verification facilitation, checking, grounding, negotiation, introducing new information, implication, and lack of confidence. Arvaniti and Atkins (2016) found that uptalk in SBrE is used to make statements, confirmation requests, and especially to hold the floor (IP-finally and before a relatively long pause) during a longer turn in a dialogue. In contrast, Armstrong et al. (2015) and Ritchart and Arvaniti (2014) report on an extensive use of plateaus to indicate floor-holding in AmE, which was rather a rare phenomenon in Arvaniti and Atkins (2016), where SBrE speakers rather used a rise (uptalk)

[56] Research on intonational meaning (as summarized in Levon 2018: 17; Brazil et al. 1980; Ladd 1983, 2008; Gussenhoven 1984, 2004; Bolinger 1986: Ch. 10; Pierrehumbert and Hirschberg 1990) hypothesized that there should be a form-meaning correspondence when it comes to intonational tunes.

or fall to indicate continuation. Levon (2016: 144) found that, among SBrE speakers, uptalk was used most frequently in the recounting of narratives, although it was also used for opinions, descriptions, and explanations, and used least to communicate facts. Barry (2008: 220) found that uptalk in London English is not only more variable in its phonetic form (compared to SoCal English), but also is more variable in the distribution of uptalk across pragmatic categories. She concludes that BrE might not have an established pragmatic meaning associated with uptalk yet, since it is considered to be a fairly recent addition to BrE intonation patterns (ibid.). According to Levon (2016: 144), uptalk is used more frequently in contexts that are interactionally more complex, i.e. while a statement including facts can be uttered within a single utterance, narratives usually require further explanation and multiple coordinated clauses in order to develop a coherent narrative. Interestingly, however, the men in Levon's (2016) study did not make such a differentiation of speech act types and the use of uptalk, which indicates that men might not use uptalk to organize the discourse in a complex way as women do. According to Levon (2016: 139), uptalk can be used in line with four principal pragmatic functions: as an epistemic marker of uncertainty, referential marker of discourse prominence, instrumental marker of turn non-finality, and affective marker of in-group solidarity. Levon (2016: 146) concludes that, while women may use uptalk more as a floor-controlling device à la Guy et al. (1986), men use it instead as an affective (politeness) device consistent with Britain's (1992) analysis of uptalk. Levon (2016: 155) adds that men deploy a referential function of uptalk to draw attention to interesting (and brand new) elements of their talk in narratives and assert their right to interpersonal involvement by doing so. Thus, women use uptalk more for instrumental purposes, but also to promote affective cohesion, while men use it referentially to obtain affective affiliation and sometimes for instrumental control (ibid.). Essentially, he sees these two functions of uptalk as "rapport-building functions" (ibid.: 157). However, he also recognizes that not all uptalk instances necessarily follow these patterns (ibid.). He explains this by saying that rising tones in general signal 'non-finality' and this allows speakers to deploy the same meaning to achieve instrumental, referential, and affective ends (ibid.: 156).

Additionally, there are many parallels between fluencemes that are used to 'buy time' in order to plan the next utterance ahead (see Götz 2013 on the functions of fluencemes) and the use of marked intonational features such as 'uptalk'. For instance, Tomlinson and Fox Tree (2011: 65) found that uptalk utterances with prolonged rising pitch are generally associated with a "forward-looking function" and utterances without such prolongations with a "backward-looking function". A similar interpretation of a combination of uptalk tunes and fluencemes (i.e. brief pauses) is interpreted by Sacks and Schegloff (1979: 19-20), as a "try marker", which are used to check the listener's understanding, and the use of the pause provides the listener with an opportunity to give the speaker a kind of confirmation (i.e. backchanneling). They further state that when no confirmation is uttered, then the "try marker" will be produced in the next utterance. Thus, these examples show that uptalk is either used in combination with pauses, as in the example of try markers (cf. Sacks and Schegloff 1979), and/or as a sort of replacement of longer silent or filled pauses, in order to gain more planning time for the utterances ahead (cf. Tomlinson and Fox Tree 2011: 61). The present study will analyze whether uptalk and fluency phenomena are used in a similar way.

This systematic variation of rises found in the studies cited in this section suggests that grouping rises as an undifferentiated category might be problematic (cf. Tyler 2019).

Many studies, as laid out in Section 2.5.3, interpret the non-canonical rises (possibly uptalk) as uncertainty on the part of NNSs. In order to find more possible explanations for the more frequent rises in non-native speech, these dramatic end-of IP and ip rises, will be analyzed in more phonetic and phonological detail. The present study will also analyze the correlation between IU length and uptalk frequency, since previous research has found that uptalk use is more frequent in longer turns, i.e. multi-clause units because the complexity of narratives also increases (Guy and Vonwiller 1989: 26), especially for women (Levon 2016: 145). The use of uptalk in native Czech, German, and Spanish is discussed in the next section.

2.8.2 Uptalk in Czech, German, and Spanish

Mentions of uptalk in German and Spanish are very few and uptalk does not seem to be a general pattern in these languages, besides a few regional varieties of these languages that have reported on such uptalk usages. Uptalk in Czech, to the best of my knowledge, has not at all been reported on in previous research, although I have personally heard it in formal and informal, male and female speech, including old and young speakers. Therefore, its potential existence (as it is used in present day English) cannot be ruled out in native Czech as of yet. However, uptalk has been reported on in different varieties of Spanish with similar uses as in English, e.g. uptalk in Spanish varieties in California (Fought 2003: 76), Puerto Rican Spanish (Armstrong 2010), and Dominican Spanish (Willis 2010: 124). In German, on the other hand, the use of uptalk has not been entirely attested, although some researchers comment on “unexpected upstep” in non-final IPs, such as Truckenbrodt (2007: 349) in his analysis of southern varieties of German (Baden-Württemberg and Austria). Furthermore, Warren (2016: 164), who conducted an extensive analysis of uptalk in English, also found a few instances of uptalk usage in German commented on by Peters (2007), who discussed the prosodic features in Franz Müntefering’s (a German politician) speeches, and by Leeman (2009), who reports uptalk in interview data in a dialect of Swiss German. Besides these few mentions of possible “uptalk” usage in German, no proof has been found that connects these uptalk usages to the ones used in English. Since the varieties of Spanish listed in the previous paragraph are spoken on the American continent, an influence by AmE, in which uptalk is very frequently used, could be assumed.

2.8.3 Uptalk in non-native speech

NNSs of English seem to be aware of the phenomenon of uptalk (e.g. Andrew 2011: 231) and seem to be adopting the feature when speaking in the L2 as well, as it was noted during the data analysis of the present study (see Section 4.4 for the results). Also other researchers have noticed NNSs using uptalk in their speech: “I also hear it quite often from young non-native speakers, who either have picked it up by ear (as it’s hardly ever taught) or perhaps use it in their native languages too” (Lindsey 2019: 108). Some studies have noted the presence of uptalk in non-native speech during the analysis of other prosodic or linguistic features (e.g. Terraschke and Holmes 2007; Grieve 2010: 204, with GerE speakers; Setter et al. 2010: 41-42, with Hong Kong English/ESL speakers; Zárate-Sández 2015, 2018; Méndez Seijas 2018, with AmE learners of Spanish), and one has made it the focus of his investigation in non-native speech (Buck 2016, with English learners of Spanish). However, most researchers report the transfer of uptalk phenomena of American learners in their L2 Spanish productions (Zárate-Sández 2015, 2018; Méndez Seijas 2018). Perhaps, one of the reasons why uptalk has not been studied as extensively in non-native speech, as it has been in NE speech, is that most studies rely on read speech as their main speaking style of investigation, especially with question intonation. The stylistic contexts in which uptalk occurs most frequently is in conversational narratives, monologues, descriptions (answers to "How?" or "Tell me about…" questions), invitations to acknowledgement from the listener, transaction openers, and answers to *wh*-questions (cf. Guy et al. 1986; Guy and Vonwiller 1989: 24-25, 34; Ladd 2008: 125). However, one of the first studies to document the use of uptalk by American learners of Spanish (Henriksen et al. 2010) analyzed different sentence-types in read speech. Another reason for a lack of evidence/research on uptalk in L2 speech might be that previous studies have written off the frequent HRTs of their speakers as uncertainty markers. To the best of my knowledge, the only researchers who explicitly analyzed uptalk in GerE speakers are Flaig and Zerbian (2016), albeit only in a small scale qualitative pilot study. They investigated whether GerE speakers (n=8; female only) use uptalk in their NG productions and/or non-native English productions in the same situation, i.e. casual speech style with the first author, with whom they were acquainted. Their preliminary analysis showed that all female GerE speakers in their study used uptalk in English but not necessarily in their NG productions. They further observed that the GerE speakers preferred a continuous steep rise, as opposed to a sudden steep rise and a continuous flat rise. They observed that HRTs were used to express uncertainty (e.g. vocabulary item) and for floor holding. Terraschke and Holmes (2007: 214) found that their GerE speakers used uptalk when communicating with New Zealanders, some of which were used on general extender expressions such as "or something" and "and stuff". According to them (ibid.: 214), these extenders and intonation are used to contribute to the interlocutor's understanding of the message and to signal solidarity. Another theory would be that NNSs might have more of a need to check whether they have been understood and are being followed (cf. the origin of uptalk and immigration in Guy and Vonwiller 1989: 33). In the few studies that observed uptalk in non-native speech in both genders, clear cases of uptalk were found in the production of both male and female L2 speakers (cf. Zárate-Sández 2015, 2018; Buck 2016). Interestingly, NSs of English have been found to use slightly more rising tunes in FDS compared to native controls (e.g. Knoll et al. 2006), which might indicate that NSs themselves might have a heightened need for comprehension checks. In turn, this might influence the learners' imitation of high tunes in their own speech when they interact with NSs.

Not much is known about uptalk in the non-native groups (CzE, GerE, and SpE) investigated in this study. We hardly know anything about the shape of uptalk, its phonetic realization, and which functions it fulfills in non-native speech. Nothing has been documented about the use of uptalk according to speaking styles in non-native speech. However, based on the literature review on uptalk in NE speech, it can be assumed that it might also be more common in narrative contexts. Therefore, the chosen genre for this study, i.e. narratives and dialogues, can be expected to contain frequent uptalk uses. Due to the frequent occurrence of uptalk in the data set analyzed in the present study, the phonological and phonetic characteristics of uptalk in non-native speech within declarative utterances (i.e. end of IPs and the end of some ips) will be investigated in more detail. The interviewer's involvement in the speech samples will also be investigated, i.e. rate of backchannels and questions asked/comments made. Unfortunately, this investigation is only limited to verbal feedback from the interviewers, since the analysis is based on utterances without visual information. Therefore, non-verbal responses to the uptalk usages of the NNSs cannot be investigated. Thus the focus of analysis will be on whether uptalk instances elicit more verbal responses than non-uptalk statements. Additionally, the interviewer gender will also be investigated in terms of uptalk usage by the non-native and native speakers. The different constellations (i.e.L1-L1, L1-L2, and L2-L2 speakers) and accommodation features will be analyzed. While the interviewers rarely use uptalk in their statements, since they mostly ask questions, I do not think that they elicit more uptalk usage in the interviewees' speech, even though uptalk usage has been described to be easily "contagious". However, the interviewers' gender might elicit more uptalk usage due to social factors, such as reducing social distance, etc. All extralinguistic variables collected for this study will be correlated with the uptalk instances, in order to find groups who are more likely to produce uptalk instances. Edelsky (1979), for instance, found with over 300 informants in a Miami, Florida university building that women produced a lot more uptalk patterns when they interacted with a female interviewer. The present study also examines the phonetic implementation of the uptalk tunes and their semantic interpretation. Given the different phonetic realizations of regional varieties of English described in Section 2.8.1, it can be assumed that learners of English with exposure to different varieties of English will also exhibit different phonetic realizations of the uptalk rises. Therefore, not only their shape but also their f0 range and duration will be measured and compared. Thus, the present study asks whether there are effects of exposure, gender, and L1 in terms of frequency of IP-final and medial rises in declarative statements and whether there are effects of exposure, gender, and L1 in terms of the phonetic realization of the rises. Further investigations will show which pragmatic functions are used with particular uptalk tunes and whether differences apply to different discourse styles (monologues and dialogues). In order to make the results comparable to previous studies, the methodologies of Armstrong et al. (2015), Levon (2018), and Arvaniti and Atkins (2016) will be adopted and extended in the present study. These methodologies involve the analysis of the f0 range (level and span), duration of rise from onset to peak, slope of rise, and uptalk instances in IP-final and ip-final positions, an AM phonological description of the tune (pitch accent + edge tones), and the pragmatic functions uptalk is used for (e.g. continuation, checking, etc.).

Since uptalk is probably non-existent in the L1s under investigation in the present study (at least in the way it is used in English), i.e. Spanish spoken in Spain, German spoken in Germany, and Czech spoken in the Czech Republic, it can be assumed that there would be no interference from their L1 in uptalk usage. Thus, if learners use uptalk in their speech this might possibly be an indication of a higher proficiency level, since their speech might be fluent enough and lexical retrieval might be more automatic. Therefore, the learners can afford to pay more attention to their f0 patterns and involve the interlocutor into the conversation by structuring their discourse through the use of uptalk. Thus, I assume that more fluent NNSs of English will use uptalk more frequently in their speech. At the same time, uptalk usage may very well be a sign of uncertainty and may, therefore, be indicative of a lower proficiency level. Whether proficiency does play a role in the use and shape of uptalk will be investigated in Section 4.4. Furthermore, it can be assumed that those learners who have been exposed to a certain variety of English (mainly AmE or AusE) will exhibit more of such uptalk usages.

2.9 Summary

The second section started out with theoretical sections, which outlined different models of intonational structure, with a focus on the AM model (Section 2.1). This model will be adopted in the present study, in order to allow for a comparison in the analysis with previous studies. Since the present study investigates learners of English from three typologically different L1s, the next Section (2.2) was dedicated to a comparison of German, Czech and Spanish prosody to English prosody in general (i.e. stress, rhythm, etc.), in order to predict which areas might pose problems to the respective L2 speakers. Then, in Section 2.3, the general problem areas, major trends and research gaps in L2 prosody were identified for the three learner groups, by giving an overview of previous studies on these specific learner groups. This was followed by the Sections 2.4-2.8, which looked into more detail on the form and function of prosody in native and non-native speech with the prosodic features of interest for the present study. The sub-sections of Sections 2.4-2.8 were organized according to the prosodic features, i.e. tones (pitch accents and edge tones) and tunes (pitch accents and edge tone configurations), intonational phrasing (ips and IPs), f0 range (f0 level and span), and uptalk, which will be analyzed and reported on in that same order in the results section.

The general picture that emerges from the studies cited in Sections 2.5-2.8 is that prosody differs substantially in native and non-native speech. Non-native speech is characterized by a reduced tone inventory (cf. Klein and Perdue 1997), shorter IUs, more interrupted IUs, a smaller phonetic extent of low tunes, and thus resulting in a narrower f0 range. However, findings of interspeaker variability and position-sensitivity of f0 range have put forth counter evidence to a narrower f0 range and NNSs thus adapting NS values. Previous research also found that learners speak more slowly in terms of speech rhythm, articulation rate, speech rate, and MLR. Fast speech, even in non-native speech, results in fewer and longer IUs, and slower speech results in more and shorter IUs. Partly as a result of more frequent IUs, a NNS of English is more likely to produce more pitch accents. Especially GerE and SpE speakers have been described to produce more level tones and replace high tones with low ones and vice versa, where the opposite would have been more appropriate. NNSs of English also tend to end their IPs with a higher or a mid level tone. Also, it has been postulated that learners of English produce low tones before high ones and complex tones follow simple ones. Most of the deviances were explained by a supposed insecurity or moderate anxiety of the learners when speaking in their L2. Previous studies on L2 prosody have shown that learners deviate in their production of prosody but it is still quite unclear what factors determine this deviation (cf. Mennen and De Leeuw 2014: 187).

Overall, it has been shown that prosody is highly context-dependent and fulfills various functions. Prosodic variation has been analyzed for how it correlates with contextual factors, such as speaker sex, age, between speech acts (questions, answers, statements, etc.), and turn position (turn-medial and turn-final), among others. This, however, has mostly been applied to NE speech. Very rarely are acoustic analyses applied to identify differences in both form and pragmatic meaning in corpora of learner English (e.g. Ramírez Verdugo 2005, 2006b; Ritchart and Arvaniti 2014). Uptalk and intonational phrasing in non-native English speech of Czech, German, and Spanish learners are virtually unexplored (see Section 2.3 for an overview). Furthermore, most studies in L2 prosody contain only very few participants, only analyze one gender (or focus on one extralinguistic variable at a time), only analyze one prosodic feature in isolation, and/or mostly analyze read speech. Especially older studies analyze only the speech of NNSs in comparison to one model NS (mostly an instructor or a very proficient public speaker) (e.g. Gutiérrez Díez 2008), even though it has been shown that there is a considerable amount of interspeaker variability within languages and dialects. Additionally, naturally occurring speech is rarely analyzed in terms of prosody (e.g. Tyler 2019 with utterance-final rises in the Santa Barbara Corpus of American English). The present study will contribute findings and interpretations for this research gap by analyzing a large group of subjects (male and female), with four different prosodic features in spontaneous speech samples and various extralinguistic variables. Besides the basic description of the three learner varieties and two native varieties of English in terms of prosody, I also intend to explain the rising end-of IU pattern in more detail. In doing so, I argue that the learner's high rises at the end do not stem solely from uncertainty but are also used as pragmatic devices. Thereby, I not only intend to describe and predict L2 prosodic patterns for CzE, GerE, and SpE speakers, but I also intend to add valuable research findings in support of the LILt hypotheses proposed by Mennen (2015).

3. Corpus data and methodology

The present section is structured first around the discussion of the approaches to L2 prosody research. Second, the corpora and tools are described. Third, the selection process of the recordings and the corresponding speaker profiles is presented. Fourth, an explanation of ToBI annotations is given. Fifth, this is followed by the annotation procedure of the present study based on ToBI and additional annotations. Sixth, the labeling procedures and inter- and intra-annotator agreements are described. The methodology outlines the dependent and independent variables of the present study, and how these were operationalized. Additionally, the statistical measures and tests are explained. Also, the research questions and hypotheses are presented. Finally, this section closes with some caveats and a summary.

3.1 A quantitative explorative approach to prosodic variation in native and non-native speech

Non-native prosody has received much attention over the last two decades, and a variety of prosodic features have been analyzed. Nevertheless, our knowledge of non-native speech is still quite limited, especially in its most natural state, i.e. in spontaneous speech. Most studies in the past as well as most current research on non-native prosody focus on one aspect of non-native prosody in read speech, with relatively small learner populations and in relatively controlled settings. A survey of 172 empirical corpus-linguistic studies carried out by Gut (2009) in L2 phonology over the past 39 years has shown that empirical research on intonation is carried out with an average number of 19.7 speakers, ranging from two to 75 (cf. Gut 2009: 44). Table 3 in Section 2.3, which includes quite recent research, offers a similar picture. In Gut's (2009) study, only studies that investigated global foreign accent and fluency reported numbers of participants exceeding 100 (43). With the exception of L1 influence, only a few studies include non-linguistic factors (such as the learning situation) in their analysis of non-native speech and the interplay among them (cf. Gut 2009: 51).

One of the main reasons for the caveats connected to previous research is the cumbersome annotation of spoken material. It is a well-known fact that the annotation of spoken corpora is a very labor-intensive task. Only one minute of speech can take up to 500 minutes of manual annotation (cf. Draxler 2008: 14). Annotation of spoken corpora becomes especially complex when it is done at the suprasegmental level, i.e. when prosodic features of speech such as intonation, rhythm, and stress are also transcribed. Therefore, phonetic corpora are quite rare. Laudable exceptions of phonetic corpora, which were mainly created to provide data for the analysis of prosody, are the LLC (Svartvik 1990), the LeaP (Gut 2010), the Aix-MARSEC[57] corpora (Hirst et al. 2009), and IvIE (Intonational Variation in English; Grabe et al. 2001). Some of these corpora were annotated semi-automatically and required many years to complete. Thus, large-scale corpus-based analyses of non-native prosodic features in spontaneous speech are extremely rare. Gut's (2009) study is one exception. A look at the LINDSEI bibliography on the Louvain website[58] reveals that studies on prosody with LINDSEI are rather scarce and center primarily around Romero-Trillo and colleagues (see,

[57] Aix-MARSEC stands for (Aix-en-Provence) Machine-Readable Spoken English Corpus.
[58] https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei-bibliography.html

e.g., Ramírez Verdugo and Romero-Trillo 2005; Romero-Trillo and Lenn 2011, Romero-Trillo and Newell 2012; Romero-Trillo 2014, 2019). The question then is whether it is possible to fully automatize the annotation process for large spoken corpora, especially those that include challenging features, such as spontaneous native and non-native speech, in order to make larger empirical studies on non-native prosody easier. During the manual annotation of the database (described in Section 3.2) for the current research project, (semi-)automatic tools, i.e. SPPAS (Bigi 2015), as well as *Praat* plug-ins (EasyAlign; Goldman 2011) were tested on the learner data. None of these tools led to the desired accuracy of annotation for IPs of my data. While these tools work well for carefully read native speech, they are not (yet) equipped to automatically label non-native spontaneous speech samples with several interruptions by interviewers in between. The resulting output files from these automatic tools led to an increased workload of manual correction, because the intonation phrase marks were off for every single utterance. Thus, a semi-automatic analysis with *Praat* combining both auditory and instrumental analysis, i.e. visual inspection of the spectrograph of the data was chosen. The exact procedure is described in Section (3.5).

The present study analyzes spontaneous speech of different native and non-native speaker groups (see Section 3.3 for an overview). As outlined in Section 3.2, the present study makes use of three different spoken corpora, more precisely large subsets of these corpora. Due to the nature of the recordings used in this study, a mainly synchronic study of the learners was undertaken, while also a pseudo-longitudinal approach was taken to study improvement after stays abroad in English-speaking countries (for a pseudo-longitudinal study see Götz and Mukherjee 2018), where learners that have not been abroad in an English-speaking country can be compared to speakers who have been abroad to investigate whether a development/improvement can be tracked. The acquisition process of the learners chosen in the present study over time could not be accounted for. The present study is a combined auditory and acoustic analysis of native and non-native English prosody. The transcripts of these corpora were used only partly, since a smaller sample of 560 words per speaker was chosen and supplemented with prosodic information. The procedure and tools used for the prosodic annotation are described in Section 3.5. A total of 225 files were analyzed, which allowed for many different quantitative measures and statistical analyses. As discussed above, a sole automatic and instrumental analysis of the corpus files was not possible, since the recordings were not made in closely controlled settings, such as read speech which entails only one speaker in a sound-proof room. The resulting *TextGrid* files from the prosodic annotation with *Praat* are a source of information different from the original corpus files. The *TextGrid* files represent a snapshot of larger recordings of each of the 225 speakers. In a traditional sense, the present study is, therefore, not entirely a corpus-based study. The resulting corpus files (*TextGrids*) are the basis for exploring the hypotheses (see Section 3.7.4). However, the new "corpus" is additionally used as a source of its own hypotheses about learner language, thus making the present study a corpus-based, as well as corpus-driven study. The present study consists mainly of a quantitative approach to productive prosody of native and non-native speakers of English. The study is, however, implicitly qualitative. Finally, the study finds itself at the intersection of a mixed methods approach to L2 prosody, i.e. it is corpus-driven, corpus-based, auditory, and instrumental.

The three interlanguages are also compared to native speech. By doing so, the study approaches an "integrated contrastive model" proposed by Granger (1996: 46), in which CA and CIA are combined. The CA from previous studies described in Section 2 helped the formulation of predictions about the interlanguage of Czech, German, and Spanish learners of English. Keeping Selinker's (1972, 1992) approach to interlanguage in mind, the present study describes the results of the prosodic analysis of the interlanguages (CzE, GerE, and SpE), initially independent of each other and then in comparison to the native control groups (NS vs. NNS) to determine deviances and similarities. Finally, the interlanguage data is compared (NNS vs. NNS). The results described in Section 4 can then be checked against the CA data, at which point an L1 influence or transfer could be investigated. This type of research can provide us with insights into SLA, because looking at data from learners with three different L1 backgrounds and four different cities (Gießen, Prague, Murcia, and Madrid) can show the developmental paths of these different learner groups and which factors help them to reach ultimate attainment in L2 prosody.

3.2 Corpora and tools

The present study is based on three interlanguage sub-corpora, as well as BrE and AmE sub-corpora as native controls. The three interlanguage corpora chosen for the analysis include learner language from different L1 backgrounds from different language families, i.e. Germanic, Romance, and Slavic. These are based on the sub-corpora of the *Louvain International Database of Spoken English Interlanguage* (LINDSEI) (Gilquin et al. 2010), i.e. LINDSEI-GE (German component; Brand and Kämmerer 2006), LINDSEI-SP (Spanish component; Pérez-Paredes 2010), and LINDSEI-CZ (Czech component; Gráf 2017). One sub-corpus for each language family has been chosen for this study in order to uncover differences and similarities between the English interlanguages of these L1s. The NS counterpart to LINDSEI is the *Louvain Corpus of Native English Conversation* (LOCNEC) (De Cock 2004), which represents the BrE component and which originated at the Université Catholique de Louvain in Belgium. LINDSEI is the first large-scale database of the spoken interlanguage of different learner populations that bridges a gap in computer Learner Corpus Linguistics (LCL) (cf. Brand and Kämmerer 2006: 127). Each national sub-corpus consists of 50 recordings with speech in monologues about one of three topics (movie or play/country/experience), free interview part, and a picture description. The monologic part is a kind of "warming-up activity, in which the learners were given a few minutes to talk about one of the three set topics" (Gilquin et al. 2010: 3). The informal interview is the main part of the of recording (ibid.: 3). Each interview lasted for about 15 minutes, which amounts to approximately 2,000 words each (ibid.: 8). LOCNEC, the comparable NS control corpus, was developed simultaneously with LINDSEI, according to the same principles, format, task design, as well as the recording and transcription procedures (cf. Gilquin et al. 2010: 65), thus making it the most suitable database for the purposes of the present study.

The Czech component (LINDSEI-CZ; Gráf 2017) was compiled at the Charles University of Prague. The 50 recordings consist of 95,904 words spoken by the learners, called "b-turns"[59] in the transcriptions. On average, an interview includes 1,919 words. The CzE speakers have had 9.9 years of English at school prior to their university education.

[59] *A-turns* are the passages uttered by the interviewer and *B-turns* are from the interviewee in the Louvain family of corpora (i.e. LINDSEI and LOCNEC).

Language proficiency level ratings were undertaken for the Czech component of LINDSEI. According to Huang et al. (2018: 10), 36 of the LINDSI-CZ learners have been rated to have an overall speaking proficiency of C1, 12 were rated at B2, and two at C2 levels.

The German component (LINDSEI-GE; Brand and Kämmerer 2006) was compiled at the University of Giessen. The 50 interviews in LINDSEI-GE amount to a total of 85,950 words spoken by the learners (cf. Gilquin et al. 2010: 25). On average, an interview includes 1,719 words produced by the learner (ibid.: 23-25). The German component was also tagged for learner errors (cf. Brand and Kämmerer 2006: 131). The German interviewees have usually had eight years of English education before starting university and are enrolled in a teaching degree, master's degree, applied languages, and business studies, or adult education program (ibid.: 132). A professional rater assessed the GerE speakers as being advanced in terms of their language proficiency, ranging from CEFR levels B2-C2 with the majority at C1 or C2 (cf. Gilquin et al. 2010: 11). Huang (2019) determined the first 17 out of 20 learners of LINDSEI-GE to be of a C1-level and the remaining three to be at a B2 level.

LINDSEI-SP (the Spanish component; Pérez-Paredes 2010) was compiled at two Spanish universities, one in Madrid (Universidad Autónoma de Madrid; n=25) and one in Murcia (Universidad de Murcia; n=25) (cf. Gilquin et al. 2010: 9-45). The entire Spanish component consists of 64,804 words (b-turns only), with an average length of 1,296 per b-turn (ibid.: 25-45). All the students were undergraduates specializing in English (ibid.: 46). The information on the learners' education and time spent abroad is unavailable for the interviews that took place at the Universidad de Murcia (ibid.: 33). A sample of the Spanish component, consisting of five speakers, had been rated as B2 and lower by a professional rater (ibid.: 11).

The 50 speakers from LOCNEC (cf. De Cock 2004) were all British students at Lancaster University in Great Britain. Most of these students were undergraduates in their first or second year but also third or fourth year, and some were also postgraduates. Most of these students were either Linguistics or English Language students, although some also studied subjects such as French, Chemistry, or Management (cf. De Cock 2004: 227). LOCNEC contains 118,564 words of spoken English produced by NSs of BrE. The interviewees were aged between eighteen and thirty years (cf. Gilquin et al. 2010: 65).

The *New South Voices* (NSV) collection was investigated as an American corpus of spontaneous speech. The NSV contains narratives, conversations and interviews in English, which are representative of the residents of Mecklenburg County, North Carolina and surrounding communities. Nevertheless, the speakers can be from other states of the US but they were included in the collection because North Carolina is now their permanent residence. The NSV-project includes over 700 audio files and transcripts of interviews, narratives, and conversations. The interviews were conducted by UNC Charlotte faculty, students, staff, as well as several community organizations. As a control corpus, I have added interview speech from the *Nationwide Speech Project* Corpus (NWSP; Clopper and Pisoni 2006), since the NSV corpus mostly includes speech from Southern speakers of the U.S.. From the NWSP Corpus I included Northern, Western, Midland, and Southern speakers. This procedure was undertaken due to the fact that the NSV was not created for phonetic analyses, and it was difficult to find suitable recordings with a decent sound quality. Following the discussion in Section 2.6.1 about the prosody and fluency of the SAmE dialect, the Midland and Northern speakers could be viewed as more representative of a General American (GA) dialect.

Since none of the corpora included a prosodic annotation of the files, the annotation was conducted with the help of the speech analysis software *Praat* (version 6.0.43, cf. Boersma and Weenink 2019) (see Section 3.5). Even though the LINDSEI, LOCNEC, NSV, and NWSP were not complied with phonetic analyses in mind, they are particularly useful for prosodic analyses in more naturally-occurring speech, which, as such, should be easier to generalize to the overall native and non-native populations. Which files of the aforementioned corpora were selected for this study is explained in the next section.

3.3 Selected recordings and speaker profiles

From these larger corpora described in Section 3.2, 45 files were selected for prosodic analysis from each sub-corpus. Thus, in total, the corpus in the present study consists of 225 files. An overview of all corpora and the number of files included in the present study can be gleaned from Table 5.

Table 5. Number of files, word count, and IP/ip duration of annotated stretches

No.	**Corpus**	**Files/Speakers**	**Words**	**IP/ip duration (min)**
1	LINDSEI-CZ (Czech component)	45	25,234	158.21/148.82
2	LINDSEI-GE (German component)	45	25,226	149.73/138.38
3	LINDSEI-SP (Spanish component)	45	24,249	160.10/140.42
4	LOCNEC (British English)	45	25,222	113.05/104.00
5	NSV & NWSP (American English)	45	25,214	122.82/111.09
Total		225	125,145	703.82

For the selection process of the 225 files, I proceeded as follows. First, I listened to all LINDSEI and LOCNEC recordings (50 recordings each). Second, I selected the most suitable recordings according to the following criteria: 1) good audio quality, 2) fluency (longer runs of speech, should not include too many obvious dysfluencies, such as constant stuttering), 3) "normal"[60] conditions (speakers should not have a cold, should not be hungover, and should not be too tired). However, a good audio quality and a certain degree of fluency were the main selection criteria. This trimming procedure resulted in the discarding of five recordings for one of the LINDSEI sub-corpora, and I opted for 45 recordings per corpus to make the corpora comparable to each other. In a third step, I selected 37 recordings from the NSV corpus (from over 700 files) and eight from dialogues from the NWSP. Due to the same design criteria of all LINDSEI sub-corpora and LOCNEC described in Section 3.2, the corpus files are highly comparable. The major similarities which make all three LINDSEI sub-corpora, the LOCNEC, and the NWSP corpus samples comparable are listed in the following:

1. 45 recordings per group
2. Same length of speech was annotated (≈ 560 tokens per speaker=1.9-5.9 minutes, on average=3.13 minutes)

[60] These cases were not made up but were actually found in the recordings. This information could mostly only be determined if the speakers clearly stated so in the recording. However, when someone has a cold and the voice is modified, i.e. sounds different, then one can hear that quite clearly.

3. Only declarative sentences, no questions (b-turns in LINDSEI and LOCNEC)
4. Non-scripted, spontaneous speech
5. Dyadic: interview situations with two speakers
6. Similar topics: country, experience, movie/theater play
7. Same age range (18-33, on average=22 years)

The NSV and NWSP generally follow these criteria as well; however, there are a few major differences that apply to some of the AmE recordings and need to be highlighted here:

1. Different recording situation in the NSV corpus: All recordings include a second person who is present at the time of the recording, although this person does not engage in the conversation as he/she does in the LINDSEI, LOCNEC, and NWSP recordings. Not even backchanneling is used (or is not audible) in most recordings.
2. Slightly different topics: In the NSV recordings, the speakers are mainly asked to speak about their favorite childhood memory. This compares most to the "experience" topic in LINDSEI and LOCNEC. Sometimes the speakers also speak about a country, but in essence the childhood memory can be about any topic.
3. Gender balance: the NE corpora (AmE only) are balanced in gender (50/50).

Due to the large selection, the NSV sub-corpus of 45 recordings chosen for this study is the most equally balanced corpus in terms of gender. Despite the differences, the corpus sample is comparable to the other samples since it includes speakers from the same age range (18-33) who are interviewed by another person. Most of the recordings even include the exact same topic as the LINDSEI and LOCNEC corpora. The topics of the conversations were comparable across the corpora, i.e. covering childhood memories, traveling experiences, etc.

After the 225 recordings were selected, according to the criteria listed above, appropriate stretches of the recordings were annotated for prosody (as described in Section 3.5). Depending on the speakers' emotional state and fluency, the most fluent and coherent part of a speaker's speech sample was selected based on my perception. Sometimes this was later on in the recording (i.e. the first 1-5 minutes of the recording was discarded), and sometimes this was right at the beginning of the recording. Whenever possible, a later stretch of the monologue (Task 1) was also annotated. The goal was to find stretches of speech that were not only fluent but also included natural f0 patterns. If one chooses the part right at the beginning of the recording, then speakers are, in general, still nervous, while at the end of recording session some speakers might exhibit potential signs of boredom (cf. Patterson 2000: 96) and, consequentially, less varied f0 patterns. Between 500-560 words were chosen as a speech sample per speaker, to make the samples comparable to most of the Murcia recordings, which were about 500 words in total, after most hesitations and interviewer stretches were deleted. In fact the LINDSEI-SP corpus is among the LINDSEI corpora with the lowest number of average words (Pérez-Paredes and Sánchez-Tornel 2015: 142). If the monologue was too short or inappropriate for the annotation procedure, a stretch of the informal interview (Task 2) would be selected. Task 3 (picture description) was left out because fluency was even lower for many learners and even NSs (see also Götz 2013). Therefore, the 2-6 minute stretches either stem from longer coherent stretches from the monologue or the informal interview (dialogue). In the informal interview, there are more intervening interviewer questions, which lead to a less coherent stretch of speech than the monologic stretches. Short question and answer pairs of 1-2 words per IP were always excluded regardless of the task production type.

Tables 6 and 7 summarize the information about the selected recordings and the respective 2-6 minute stretches from the native and non-native corpora with average measures on task length.

Table 6. Information about the recordings (learners). NA cells indicate information that was not available for the corpus

	LINDSEI-CZ	**LINDSEI-GE**	**LINDSEI-SP**
Date of recording (year)	2012-2014	2004	2000-2002
Topic of recordings	country=21 entertainment=13 experience=11	country=33 entertainment=9 experience=3	country=27 entertainment=10 experience=8
Speaking style	mono=9 dia=11 both=25	mono=8 dia=10 both=27	mono=5 dia=10 both=30
Total task lengths (in minutes)	Task 1=91.1 Task 2=166.53	Task 1=92.45 Task 2=170.28	Task 1=86.93 Task 2=153.82 Madrid: 45.67 & 85.93 Murcia: 41.27 & 67.88
average sample length (unpruned)	5.71 min	5.83	Madrid=5.27 Murcia=5.45
Familiarity with interviewer	Familiar=23 Vaguely familiar=22	NA	Vaguely familiar=20 Unfamiliar=25

The Spanish and German corpus come from the same time periods (time of recording), while the Czech corpus is the most recent addition. “Country” is the most popular choice of topic among all three corpora, followed by the movie (entertainment),[61] and personal experiences. The present study distinguished between three speaking styles for the selected samples, i.e. monologue meaning only pure speech by the interviewee, dialogue meaning a conversation with the interviewer and interviewee involving question and answer pairs, and a mixture of both speaking styles, i.e. “both”. Even though the LINDSEI annotations of each transcript provide a distinction between the three tasks (Task 1=monologue, Task 2=dialogue, Task 3=picture description), the present thesis determined speaking style of each recording with the help of the following factors:

Monologues: a sample was labeled as purely monologic if the interviewee speaks on their own, with only 1-2 minor interruptions by the interviewer, which do not interrupt the flow of speech drastically.

Dialogues: a sample was labeled as purely dialogic if there was a clear exchange of question and answer sessions between the interviewer and interviewee.

[61] While the original LINDSEI label for this topic is “film”, I labeled it “entertainment” to include samples where interviewees speak about a book, a film, or a theather play.

Both: This speaking style includes both monologic and dialogic stretches. Most recordings include both speaking styles because most monologues in LINDSEI are rather short and then go over to the dialogic part.

As with the topics, the speaking styles are also almost equally distributed among the three learner groups, with “both” speaking styles characterizing most of the recordings, then dialogues, and then monologues. The familiarity with the interviewer is available only for the Czech and Spanish corpus. Table 7 gives an overview of the speaker profiles of the learners.

Table 7. Speaker profiles (learners)

	LINDSEI-CZ	**LINDSEI-GE**	**LINDSEI-SP**
Age	20-27 years, mean=22.44, median=22	21-33 years, mean=24.58 , median=24	19-29 years, mean=21.6, median=20
Gender	male=6 female=39	male=13 female=32	male=11 female=34
Home language 1	Czech	German	Spanish
Home language 2	Finnish (1), Other (1), English (1)[62]	English (1)	French (1), Catalan (1), English (1)
Mean length of years of studying English at school prior to university	10.1	8.7	only Madrid sub-corpus: 8.9
Mean length of years of studying English at university	3.4	3.7	only Madrid sub-corpus: 3.4
No stay abroad	8 learners	12 learners	6 learners (maybe more)
Mean length of stay in an English-speaking country (in months)	0.25-168 9.6	1-60 9.1	0.25-24 only Madrid sub-corpus and part of Murcia sub-corpus: 5.1
English speaking countries travelled to	GB=8 USA=7 Ireland=4 Scotland=2 USA/GB=1 Australia=1 India=1 Ireland/GB=1 NA=12	GB=16 USA=15 Australia=2	GB=12 Ireland=7 Ireland/GB=3 USA/GB=2 GB/Canada=1 GB/Ireland=1 GB/Scotland=1 GB/USA=1 USA=1 Ireland/GB/USA=1 Scotland=1 USA/Ireland=1 Wales=1 NA=6
L3	German (25), French (11), Spanish (6), Dutch (1), Italian (1), Russian (1)	French (21), Spanish (12), Italian (4), Other (4), Russian (1)	French (14), German (6), Arabic (1), Russian (1), Swedish (1), information NA (18)

[62] Even though one learner from the LINDSEI-CZ corpus indicated to use English at home 5% of the time, the learner was included because the main L1 used at home is Czech.

What all learners in the present study have in common is that they all are university students, with mostly one main home language (few exceptions with another home language) and with comparable lengths of years of English at school and at university. For most learners a semester abroad for all three universities is compulsory. However, eight learners from LINDSEI-CZ, 12 from the German corpus and six from the Spanish corpus have not spent any time abroad in an English-speaking country at the time of the recording. Those learners who have had a stay abroad in an English-speaking country stayed there on average for 9.6 months (Czech), 9.1 months (German), and the Spanish learners had the shortest stays with 5.1 months on average (data only available for the Madrid corpus and part of Murcia, NA=8). In the present corpus, there are 18 learners who have been on long stays abroad of (13-168 months), 83 learners have been on short stays abroad of 0.25 months to 12 months, 26 learners have not been an in an English-speaking country, and for eight Spanish learners there is no information available for stays abroad. The distinction between short and long stays abroad is based on Götz and Mukherjee's (2018: 51) study of a subset of the data set (LINDSEI-GE) and the stay abroad variable, in order to make the results comparable, i.e. anything under 12 months is considered a short stay abroad and anything over 12 months a long stay. Most learners in all three corpora have been abroad in Great Britain and the USA. Some learners, especially the Spanish ones, have been abroad in two other English speaking countries. However, the stays of these study abroad, work and travel, or aupair programs were quite short for the Spanish learners. The most popular third languages are German and French for the CzE, French and Spanish for the GerE, and French and German for the SpE speakers.

The native counterparts to the LINDSEI sub-corpora (CZ, GE and SP) in the present data set are the LOCNEC (BrE) subset of 45 speakers and parts of the NSV and the NWSP corpora representing AmE. More detailed information about the NS groups and the respective recordings is listed in Table 8 and 9.

Table 8. Information about the recordings (native speakers)

	LOCNEC	**NSV/ NWSP**
Date of recording (year)	1995-1997	1998-2004 (NWSP only 2003)
Topics of recordings	experience=6 entertainment=12, country=21, self=6	experience=22 entertainment=1, country=1, self=13 / 8 (NWSP)
Speaking style	mono=7 dia=29, both=9	mono=24 dia=20 (8 from NWSP), both=1
Total task lengths (in minutes)	Task 1=68.88 min Task 2=155.9 min	Total length of monologues: 76.9 min Total length of dialogues: 22.68 min (NSV), 32.4 min (NSWP) Total length of mixed: 3.9 min
Average sample length (unpruned)	5 min per recording 1.53 min (Task 1) 3.47 min (Task 2)	NSV=3.48 min NWSP=4.05 min Avg. length of monologues: 3.33 min[63] Avg. length of dialogues: 3.37min (NSV), 4.05 min (NWSP)
Familiarity with interviewer	Unfamiliar=37 Vaguely familiar=8	NA

[63] The lengths of the AmE corpus are not comparable to the lengths of the other four corpora, because in the AmE corpus each recording mostly consists of one speaking style only (only one recording contains features of both speaking styles). The other corpora can contain recordings that have Task 1 and Task 2 components and, therefore, the time might be based on a smaller sample per speaker.

As Table 8 demonstrates, dialogues in LOCNEC are just as frequent as monologues in the AmE corpus (including both NSV and NWSP recordings). Additional biographic information on the NSs is provided in Table 9.

Table 9. Speaker profiles (native speakers)

	LOCNEC	**NSV & NWSP**
Age	18-30, mean =21.4, median=20	18-33, mean=22.73, median=23, NA=3
Gender	male=20 female=25	male=20 female=25
Home language 1	all English	all English
Home language 2	all English, except 1 French (BE004)	NA
Foreign Languages	all monolingual English except: French (4), German (2), Italian (1), Chinese (1)	NA

Both native corpora consist of a quite balanced amount of male (n=20) and female (n=25) speakers. The age range is also very similar, even though the BrE speakers are on average a bit younger. 16 speakers of the American corpus are students; other speakers have occupations that include: computer technician (1), marketing (1), day care provider (1), communications (1), insurance analyst (1), insurance underwriter (1) psychologist (1), quality control (1), restaurant owner (1), sales (4), teacher (1), waitress (1), teambuilding facilitator (1), web designer (1), and finance manager (2). The speakers from the NSV corpus have a different socioeconomic background than the other groups, 19 out of 37 speakers have a professional status and are not students. However, the NSV speakers are all in the same age group as the LOCNEC and LINDSEI speakers, which leads to the assumption that they quite recently finished their education and are new to the workforce. In the NSV corpus, there are additionally seven NAs concerning occupation. All eight speakers from the NWSP corpus are students.

The choice of corpora resulted in the four possible constellations of interaction between two interlocutors, listed below.

- L1-L1: native interviewer and native interviewee (e.g. NSV)
- L2-L2: non-native interviewer and non-native interviewee (e.g. LINDSEI-CZ)
- L2-L1: non-native interviewer and native interviewee (e.g. LOCNEC)
- L1-L2: native interviewer and non-native interviewee (e.g. LINDSEI-GE, CZ, SP)

Further information on the interviewers can be found in Table 10. The corpus thus includes either narratives, which are monologic (depending, however, on the interviewer), and conversational interviews. Problems with conversational interviews are that they are of varying length of speaker turns. Many learners rely on the interviewer too much to ask questions and do not engage in the conversation extensively. Another problem is that learners scaffold on the interviewer's contributions.

Table 10. The LINDSEI and LOCNEC interviewers

Corpus	Interviewer 1	Interviewer II	Interviewer III	Interviewer IV	Interviewer V
LOCNEC	1 non-native female Belgian L1 (all 45 interviews) (labeled as BgE)				
LINDSEI-CZ	1 BrE female (22 interviews BrE_3)	1 native Czech male (23 interviews) (labeled as CzE)			
LINDSEI-SP	1 AmE female (18) from Madrid (labeled as AmE_3)	1 AusE female (7) from Madrid (labeled as AusE_2)	1 BrE (from Wales) female (20) all from Murcia (labeled as WaE)		
LINDSEI-GE	1 AmE female (16) (labeled as AmE_1)	1 British/German bilingual female (17) (labeled as BrE_1)	1 AusE male speaker (3) (labeled as AusE)	Female 1 (4) - BrE (labeled as BrE_2)	Female 2 (5) AmE (labeled as AmE_2)

The information concerning speaker status (native vs. non-native) and the varieties of English spoken by the interviewers are important for the analysis of prosodic entrainment and the multifactorial analysis in general. In total there are only two NNSs of English in the role of interviewers, i.e. in LOCNEC and LINDSEI-CZ; all others are interviewed by NSs ("monolingual" and bilingual speakers). All 45 LOCNEC files were interviewed by the same interviewer, i.e. a non-native female speaker from Belgium. 37 interviewees were completely unfamiliar with the interviewer and eight stated that they are vaguely familiar with her. The Czech sub-corpus was evenly split up between two interviewers, a female BrE interviewer who conducted 22 interviews and a non-native male speaker of English (the other BrE interviewer's husband) from the Czech Republic with 23 interviews. All 22 interviewees were vaguely familiar with the female interviewer and all 23 interviewees were familiar with the male interviewer. The Spanish sub-corpus included three female interviewers representing three different varieties of English. The interviews in the sub-corpus from Madrid, including 25 interviewees, were conducted by a female AmE speaker (18 interviews) and an AusE speaker conducted the remaining seven interviews. All 25 students from Madrid were unfamiliar with the two interviewees. The interviews from the Murcia sub-corpus (including 20 interviewees who were all vaguely familiar with the interviewer), were conducted by a Welsh speaker of English. Finally, the German sub-corpus exhibited the largest variety in terms of interviewers (four female interviewers and one male), involving three different varieties of English (AmE, BrE, and AuSE). The majority of the interviews were conducted by the BrE/German bilingual (17 interviews) and the first AmE speaker (16 interviews). Unfortunately, the familiarity with the interviewees was not documented for LINDSEI-GE and it is, therefore, unknown whether the interviewees knew that the two main BrE and AmE interviewers were highly proficient in German as well.[64] As becomes apparent from this overview, male-male conversations are quite rarely found in the corpus of the present study, since only 31 out of 225 interviews were conducted by males (only in LINDSEI-CZ and LINDSEI-GE), and only seven out of 225 involved male-male conversations (two from LINDSEI-CZ, one from LINDSEI-GE, and four from NSV). In total, there are 24 male interviewer interactions with female interviewees and 65 female interviewer interactions with

[64] I know this information because I know the two of the interviewers personally.

male interviewees. The remaining 128 recordings out of 225 involve female-female interactions (56% of all interactions). Thus, the corpus consists of dyadic mixed gender, as well as female-only, and male-only conversations.

3.4 ToBI notations

In order to analyze English intonation, there have been attempts to develop language-independent transcription systems, such as Halliday's English intonation system (1970), *International Transcription System for Intonation* (INTSINT; Hirst and Di Cristo 1998), Pierrehumbert's intonation model (1980), the ToBI (Tone and Break Indices) transcription system (Silverman et al 1992), or Ladd's AM theory (1996), and many more. The ToBI transcription system, which I adopt for the present study, is based on AM theory and MAE_ToBI was developed by Silverman et al. (1992), while guidelines and training material have been provided by Beckman and Ayers (1997). ToBI is the preferred transcription system in this study because it allows for a category-based interpretation of intonation which is compatible with leading theories of SLA (for a discussion of these models regarding prosody, see Section 2.3) such as the SLM (Flege 1995) and PAM (Best 1995) (cf. Jilka 2007: 81). It is important to point out that ToBI is a *phonological* labeling system based on the AM model of English intonation. Many researchers (used to) think that ToBI is *an International Phonetic Alphabet* (IPA) for prosody. Currently, an IPA version of ToBI, called *International Prosodic Alphabet* (IPrA), is being discussed in the speech prosody community, and first arguments for a broad phonetic notation system in addition to the existing phonological one (ToBI) has been put forward (cf. Hualde and Prieto 2016). Hualde and Prieto (2016) argue for two levels of prosodic representation, a broad phonetic and phonological level based on the same set of symbols (ToBI phonemic labels). A need for such an approach has been for instance applied to previous prosodic research, e.g. Clopper and Smiljanic (2011) who adopted a phonetic approach to describing change in f0 with the existing ToBI labels, although these were originally intended to capture phonological properties of intonation systems. A similar approach is proposed by Face (2005) for the distinction of declaratives and interrogatives in Castilian Spanish. When employing a ToBI transcription, one has to adopt a combined approach of instrumental analysis of the automatic tracking of f0 with a speech analysis program such as *Praat* and an auditory verification of the speech signal.

According to Beckman and Ayers (1997), a ToBI transcription consists minimally of three things: a speech recording, an electronic or paper record of the f0 contour (which can be obtained from speech analysis programs like *Praat*), and four tiers with symbolic labels. The four tiers ToBI proposes are the following:

1. A tone tier
2. An orthographic tier
3. A break index tier
4. A miscellaneous tier

An example of a speech signal annotated with ToBI notations is shown and explained in Section 3.5. The final annotation included seven additional tiers in addition to the four basic ToBI tiers (see Section 3.5 for the annotation procedure and *Praat* screenshots). The inventory of pitch events possible within the ToBI framework on the tone tier are based on the AM model of intonation. Phrase accents and edge tones can only consist of either H or L target tones. These tones can be combined into composite pitch accents, L+H and H+L, with the plus symbol (+) linking the two tones. Tones are determined by both an interpretation of the f0 line and an auditory verification. Just as described in Section 2.1, pitch accents only occur within prominent syllables. Each IP and ip has to have at least one pitch accent on stressed syllables, which is marked by an asterisk (*). The exclamation mark (!) is used to mark downstepped accents, e.g. !H* is a high pitch accent following a previous H* pitch accent but with a lower f0 height than the previous one, also called utterance declination. The pitch accent is always followed by a phrase accent marked by a hyphen (-), which stretches to the end of ips. The final boundary tone falls exactly on the IP boundary of stressed or unstressed syllables and is marked by the symbol %. Any change in f0 before pitch accents and in between are not transcribed; this is reserved for f0 minima or maxima only (cf. Grice and Baumann 2007: 45). Since H and L are relative terms, a 'low' boundary need not indicate a fall, but can show a flat, non-rising f0 at the end of a phrase (cf. Warren 2016: 12). For a more detailed discussion of ToBI conventions, the reader is referred to Beckman and Ayers' (1997) ToBI Labeling Guide.

Having primarily been developed for AmE, ToBI is nevertheless flexible enough to be adopted for other languages as well, while retaining the same analytical framework (cf. Jun 2005). Since ToBI has been adapted for many other languages, the original ToBI transcription system for AmE is now called MAE-ToBI (Beckman et al. 2005; Ladd 2008: 105). There is a transcription system for Standard German, GToBI, which is based on speech data mainly from Northern German speakers (cf. Grice and Baumann 2002, Grice et al. 2005 for an overview). The original Sp-ToBI transcription for Spanish has been proposed by Beckman et al. (2002) and modified in Face and Prieto (2007), as well as in Estebas-Vilaplana and Prieto (2008, 2010). No generally agreed upon ToBI notation system for Czech or non-native speech has been developed yet (see, however, Pešková 2017 for a first proposal). For a comparison of these different ToBI systems, see Section 2.5.2.

Despite a few shortcomings, ToBI was chosen as the most adequate notation system for the present study, because ToBI is the most widely used annotation system for intonational features. Therefore, it allows for a direct comparison between previous studies, while additionally making direct cross-dialectal comparisons and mapping of tunes possible. In the present study a slightly modified version of ToBI was used and all transcription systems for the different languages of this study were consulted. How this was accomplished is described in the following section (see also Section 3.7.1.1 for the combined tone inventory). Modified versions of ToBI are common practice, especially when dealing with the speech of NNSs of English (e.g. Arvaniti and Baltazani 2005; Gut 2009; Mennen et al. 2012). These studies show that the ToBI notation system is flexible enough to account for non-native speech as well.

3.5 Procedure of annotation

The selected audio files, as outlined in Section 3.3, were annotated manually using the speech analysis software *Praat* (Boersma and Weenink 2019). This involved repeated listening to the acoustic signal, and visual inspection of the spectrograph, and the f0 track. For the labeling of the prosodic features, a slightly modified version of ToBI (Silverman et al. 1992) annotation system was used. The background and details of the ToBI annotation system have been described in Sections 2.1 and 3.4. Since the present study is based on English speech spoken by different native and non-native speaker groups, MAE-ToBI (Beckman et al. 2005) was chosen as the major source for the merged tone inventory and the corresponding annotation labels. Which tones were included in the tone inventory, in order to analyze both native and non-native English, is explained in Section 3.7.1.1. As mentioned in Section 3.4, ToBI annotation consists of at least four parallel tiers, i.e. an orthographic tier, a tone tier, a break-index tier, and a miscellaneous tier. As can be seen in Figure 2, the 225 files involved the annotation of 11 manually added tiers. Each of the 225 files underwent a certain labeling procedure and annotation checks, which are described in Section 3.6. In the following, I will briefly describe each annotation tier. A detailed annotation manual with a comparison of the original ToBI annotation can be found in Appendix 1.

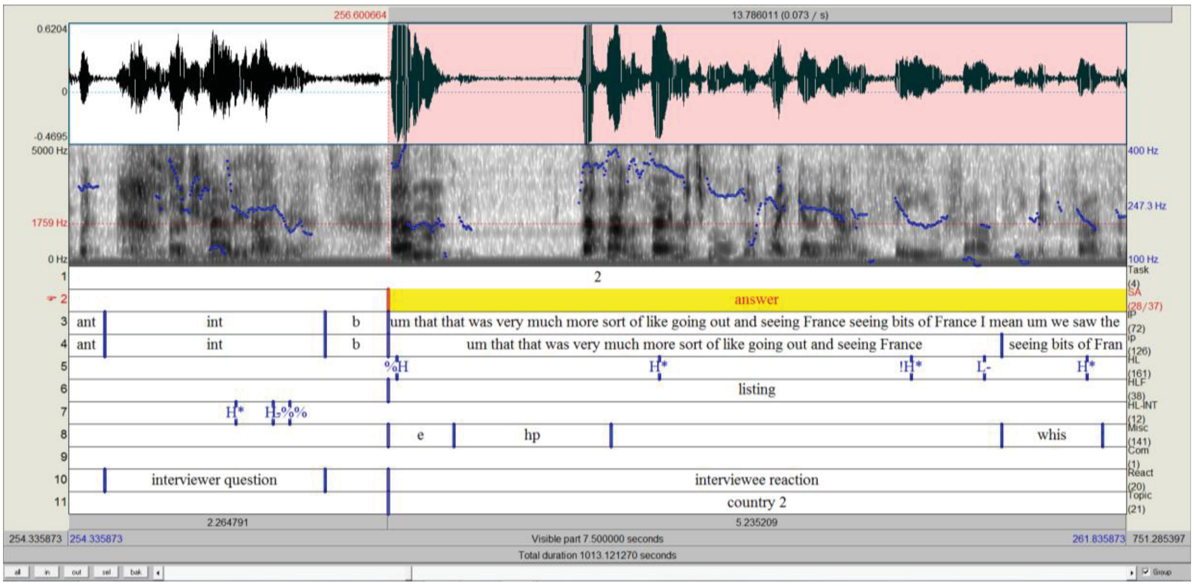


Figure 2. *Praat* screenshot of all annotation tiers (file BE018)

Figure 2 illustrates the annotation of the corpora with the help of the wave form, spectrogram, and the f0 track. From top to bottom, the tiers are called task tier, speech act (SA) tier, intonation phrase (IP) tier, intermediate phrase tier (ip), tone tier (HL), pragmatic function of tones (HLF) tier, interviewer tone tier (HL-INT), miscellaneous tier (Misc), a comment tier (Com), an interaction tier (React), and a topic tier. All tiers were annotated as interval tiers, except the two tone tier (HL and HL-INT), which have been annotated as point tiers.

The task tier (tier one) indicates whether the annotated speech sample took place in Task 1 (monologue) or Task 2 (informal interview). This is an important distinction in order to be able to analyze task effects on the prosodic features.

The data was also coded for speech acts. On the speech act (SA) tier (tier 2), two labels were annotated. Each IP was coded for utterance type, i.e. whether it functioned as a statement or answer. Even though the interviewees occasionally ask questions, especially declarative questions, only declarative statements were annotated. All questions, including statement questions, echo questions, as well as tag questions were removed. Utterances were considered questions if they could possibly be answered with a yes/no response, or if they were *wh*-questions, and syntactically marked with a *wh*-word. Even if only one ip contained a question or question-like structure, the entire IP was discarded. The declarative utterances were further distinguished between answers and statements. The distinction between answers and statements by the interviewees was made as follows:

- Answers (A): Speech samples from Task 2 (the informal interview) would most commonly have speech acts labeled as “answers”. Once an interviewer asks a question, all subsequent utterances from the interviewee are labeled as “answers”. However, there are sometimes questions posed by the interviewer that resemble more statements but are in fact questions. In these cases, the interviewee utterances were labeled as “answers”.

- Statements (S): Utterances were classified as “statements” that are usually part of Task 1 (the monologue), and that are mostly “free of any” influence from the interviewer. When the interviewer makes a statement, the following utterances by the interviewee are labeled as “statement”. Once a question has been posed by the interviewer, the subsequent utterances are usually labeled as “answers”, however, when the interviewee starts off with a new topic all by him-/herself after a while, these utterances are labeled as “statements” again.

On the IU tiers, utterances were segmented into intermediate (ip) and intonation phrases (IP) on tiers three and four (cf. Figure 2). The separation into IPs and ips corresponds to the ToBI break indices of three and four. Word and syllable boundaries were not annotated. The segmentation process involved various steps. Every IU received an initial and a final landmark, where a visual and auditory inspection of the spectrograph and the f0 line indicated a break by means of a pause (silent or filled), final syllable-lengthening, a change in f0 level, or a combination of these features. Most IP boundaries were identified by longer pauses and a final lowering of f0. Also, the context was considered, i.e. if an IP had been interrupted by an interviewer before the interviewee could end it, it was discarded from the annotation process. Often IPs coincided with clause boundaries and ips with phrase boundaries in syntactic terms. Only complete IPs of a certain length (at least five words) were annotated. Shorter question and answer pairs were excluded. Syntactic completeness did not play a role and syntactically incomplete utterances were included in the analysis. Thus, many IPs end with a conjunction such as *and* or *so* plus a low boundary tone. After the landmarks for phrase boundaries have been set, the orthographic transcriptions were added in the respective intervals. Filled pauses such as *um*, *er*, *mm* were counted as words in the IPs, since they also fulfilled a prosodic function (i.e. they are potential targets as tone bearers) and contribute to the overall sentence rhythm, i.e. the nucleus can also fall on such fillers and they can be produced with any intonation, mostly, however, they are produced with a level f0 and they can be separated from other ips as separate phrases (Clark and Fox Tree 2002: 100). The spellings *er* and *um* for filled pauses are usually the BrE spellings but they were used for both AmE and BrE transcriptions. A differentiation between different types of filled pauses was not undertaken (see Clark and Fox Tree 2002: 75).

On the tone tier (HL; tier five), prominent syllables and phrase boundaries were first identified and a point was placed on the point tier in *Praat*. These points were then filled with ToBI-labels in a second step. Only two basic tone labels are distinguished, i.e. high (H) and low (L). Following ToBI guidelines, each tone received a diacritic according to their position within an IP. There are three different positions a tone can be placed. A boundary tone, for instance, appears at the very beginning or end of an entire IP and is marked with a %, pitch accents receives a * and they can appear anywhere in the IP after an initial boundary tone and before the first phrase accent. All phrase accents receive a - and they can appear on accented or unaccented syllables towards the end of IUs. In accordance with ToBI notations, bitonal combinations were only allowed for pitch accents, which are marked with a +, e.g. H+L*, which is a low tone target on the accented syllable immediately preceded by a high tone. For all high tones a distinction was made between the f0 height of high tones in relation to the preceding or the following high tones, i.e. upstepped (marked with a ^) and downstepped (marked with a !) high tones. An upstepped tone is markedly higher than the preceding high tone(s) within an ip, while an upstepped tone can be a slightly or markedly lower high tone than the preceding ones within ips. The presence or absence of these tones was determined on the basis of the ToBI criteria outlined in Section 3.4. Table 12 in Section 3.7.1.1 provides a list of all possible tones and combinations used in the present study, while the manual in Appendix 1 offers a detailed explanation of the annotation procedure of tones.

In addition to the tonal choice, the corpus in the present study has also been labeled with interactional acts indicating the pragmatic effect of turns. The pragmatic function of the utterances was determined by the discourse context but mainly the tune was considered to make decisions on the overall pragmatic function of an ip. Each pitch accent, phrase accent, and boundary tone contributed to an overall meaning or function of each ip or whole IPs. Within each IP, ips can change their pragmatic function. Often more than one pragmatic function can apply. Generally, boundaries between pragmatic functions are fuzzy and they cannot always be clearly distinguished. Especially since it cannot be determined with absolute certainty which pragmatic function the interviewee intended for each respective ip, it can only be inferred from the discourse in the given speech sample. Above all, with NNSs with a lower proficiency level, and/or the use of misleading tunes, can make it even more difficult to determine which pragmatic function was intended. However, in order to make the data analyzable, the main pragmatic category that applies to each ip was chosen. Based on the discourse situations in all of the recordings, the following pragmatic functions (tier six) were identified and annotated for every ip based on their f0. The pragmatic functions are ordered in their frequency of occurrence, starting with the most frequent ones in the present data set (a detailed explanation of the pragmatic functions can be found in the annotation manual in Appendix 1): Continuation, handing over turn, listing, finality, checking, insecurity, and emphasis.

Following the pragmatic function tier of the interviewee utterances, the interviewers tone labels were annotated on another point tier (tier seven). The interviewer tones that stem from either questions or statements immediately before interviewee turns were annotated in order to make the analysis of prosodic entrainment possible.

The miscellaneous (Misc, tier eight) tier includes various paralinguistic and non-linguistic information that are not part of the prosody of an utterance, (e.g. coughs, laughs, etc.). A detailed list of all these features is given in the annotation manual in Appendix 1. The miscellaneous tier annotations will help to put all intonational events into perspective and to explain certain deviances in a more detailed way.

The comment tier on tier nine is reserved for any type of comments from annotators or additional (non-)speech events can be added, which are occurring simultaneously.

The interviewer and interviewee interaction tier (tier ten) helps to determine whether the interviewee speech is influenced by the interviewer, i.e. prosodic entrainment effects on the tonal patterns and intonational phrasing. On this tier, only three possible options can be selected: 1) "interviewer statement", which is a statement made by an interviewer that can be commented on by the interviewee or not; 2) "interviewer question" includes any type of questions asked by the interviewer directly before interviewee turns; 3) "interviewee reaction" is then annotated when the first interviewee turn has directly affected the interviewer question or statement. If the interviewee ignores the interviewer question or statement than the "interviewee reaction" annotation is not added.

On the topic shift tier (tier eleven), different topics talked about in the speech samples are annotated. In this manner, topic shifts and changes in intonation can be analyzed and it can be determined how coherent the annotated stretches are. The list of topics identified in the present data set is given in the annotation manual in Appendix 1.

3.6 Labeling procedures, annotators, and intra- and inter-annotator agreement

The annotation procedure described in the previous section involved seven different annotators who were all student assistants, mainly undergraduate students in English Linguistics, except for the author of the present study. All student assistants were German NSs (L2 English) and one was an AmE-German bilingual. The annotators were regularly trained with various workshops on identifying intonational features on a weekly basis for a period of two years (2015-2017). In the context of this study, the skill of *phonological awareness* is defined as the ability to detect, distinguish between, segment and manipulate speech sounds (Oakhill and Kyle 2000: 152). Generally, foreign language learners, such as the ones who annotated the recording for the present study, do not possess a phonological awareness necessary or suited for the present study. Furthermore, during their course of studies in Linguistics, prosody is generally not an extensive part of their studies. For these reasons, the annotators received regular training for two years.

Each one of the 225 recordings was subject to the same procedure, i.e. first annotation by annotator one (A0), annotation check by annotator two (A1), and final annotation check by annotator three (A2). The second annotator in the A1-step, duplicated the tiers and checked for accuracy. If the second annotator agreed with the annotations made in A0, she would leave these stretches blank. Annotator three would then check the A1-annotations in the A2-step and make the final decision on tones and intonation phrase boundaries. Problematic transcriptions were marked on the comments tier and those transcriptions were examined by all annotators in the next meeting. In cases of disagreement, the data was discussed until an agreement could be found. In phonetic and phonological studies in Clinical Linguistics, this method is called a “consensus method”, where two or more coders discuss a certain feature until a collaborative final version is reached (Shriberg et al. 1984). The annotation roles (A1, A2 labelers) alternated between the seven different annotators. However, the A3 was mostly reserved for me, unless I was involved in the A1 or A2 phase of an annotated speech sample. Once these three labeling steps were completed, the annotations were crosschecked for correctness and consistency according to the labeling manual. In order to assess the reliability of the labeling system (see manual in Appendix 1), inter- and intra-rater agreements (Krippendorff 2004) were calculated for a small subset of the entire corpus. Krippendorff’s (2004: 214-216) offers three different kinds of reliability tests: (1) Stability: carrying out the coding, then redoing it at a later stage, (2) Reproducibility: different coders code the same data, (3) Accuracy: the coders are able to code in the same way as a known standard. In order to achieve that, a learner sample recording was randomly selected and each annotator labeled the same recording (inter-annotator agreement). Also the annotators were asked to annotate a recording which they have annotated at the beginning of the annotation training and one after at least 6-12 months of training (intra-annotator agreement). Since the students joined the project at different times and for different periods, I only included the three principal annotators in the inter- and intra-annotator agreement calculations. The reason for that being that these three annotators were involved in almost every single file of the corpus, especially in the final stages of the annotation process.

The presence of certain cues has been shown to aid the segmentation on the prosodic level. Brugos et al. (2018) investigated the role of six timing, pitch, and voice quality cues in signaling boundary events when using a ToBI-labeled AmE corpus for annotation. They found that the placement of IU boundaries increases with the number of cues present at the boundaries. Their findings also suggest that labeler uncertainty increases for boundaries that include fewer cues (Brugos et al. 2018: 245). Thus, the more timing, pitch, and voice quality cues are present the more they coincide with breaks annotated by ToBI labelers. Another interesting factor that Brugos et al. (2018: 248) mention is that the absence and/or weakness of one cue to a given category may be compensated for by the presence and/or strength of a different cue. Which means that if we have one very strong cue, this might compensate for many weaker ones. Therefore, the labelers were trained to pay attention to these cues to IU breaks and to weight on their number and strength. Only the IU tiers, tone tiers, and miscellaneous tiers have been annotated and corrected by three labelers each. The other tiers, e.g. pragmatic function, speech act tier, etc., have been added later by the author of the study. These steps were necessary because previous studies have shown that reliability of the annotations is influenced heavily by the complexity of the annotation task (Gut 2009: 72; Gut 2012). In Gut’s (2009) study, only the segmental, word, and pitch tiers reached perfect inter- and intra-annotator agreements, while syllable and tone tier annotations showed low reliabilities. Gut (2009: 73) found that the higher the number of different tiers and annotation schemes is, the lower the agreement will be between annotators. Therefore, I deemed it appropriate to only let the annotators (including myself) focus on four tiers (IPs, ips, tone tier, and miscellaneous tier) at a time. They were also advised in the weekly meetings to annotate the IPs first, then the ips and then the place a point on the point tier where they heard a change in f0 level on a prominent syllable. As a final step, the labelers were advised to fill out the tone labels for each point tier. This is a procedure I also followed myself.

The functions ckappa and lkappa (from the library psy) were used to compute the kappa coefficient and test how well two or more raters conform in their judgments of stimuli (cf. Gries 2013: 246). Cohen's kappa (Cohen 1960) was used to calculate inter-rater agreements. The average kappa value across annotators was 0.71. Inter-annotator agreements were calculated for the IP and ip tiers as well as the tone labels. The transcriptions on the tone tier reached kappa values of 0.62, indicating only moderate agreement. The mean kappa for the IP tier was 0.95, indicating almost perfect agreement among annotators. Lesser agreement was found on the ip tier with a kappa value of 0.73. This finding is in line with previous research, where raters agreed on the placement of IP boundaries more often than on the placement of ip boundaries. As Grice et al. (2017: 90) write, "establishing the presence or absence of a pitch accent is relatively straightforward, distinguishing between different types of pitch accents is not". To compare that with previous research, for the labeling of tones, Ulbrich (2012: 7) found that 78% of annotations were consistent across annotators. She found the highest variability in the annotation of rising patterns (17%). In Ulbrich's (2012) project, the annotators collaboratively revisited the problematic f0 patterns and decided on the label through a joint visual examination of the f0 lines. This procedure was also employed in the present study. After a student send in a finalized *TextGrid*, we decided on the tone labels collaboratively wherever the annotator left a question mark. Speer et al. (2011: 47) report an over 90% agreement rate among five transcribers on the location of tones (i.e. pitch and phrase accents, and boundary tones). However, when it comes to the type of tone only 60% for pitch accents and 83% for boundary tones was observed. This highlights the fact that even if prosodic annotation is based on a standard like ToBI, there is still a large amount of subjectivity in the decisions that are being made. Even seemingly 'simple decisions' such as the distinction between high and low are not as easy to make.

Due to the low inter-rater agreements, all annotations were checked by the author and final corrections were made to make the prosodic labels more uniform across all files. Nevertheless, the experience of involving student assistants in the annotation process was a valuable decision. The constant stimulation and challenging impulses from different viewpoints on prosodic variation and annotation aided me in arriving at sound decisions concerning the annotation of native and non-native speech.

3.7 Methodology

The present study extends the commonly analyzed prosodic features in previous research, i.e. f0 range (in native and non-native speech) and uptalk (in native speech) (see Mennen et al. 2012, 2014; Levon 2016, 2018) by including several dependent and independent variables, linguistic as well as extralinguistic ones. The present section describes the dependent and independent variables in more detail and detailed research questions are stated.

3.7.1 Productive prosody: Quantitative analysis and dependent variables

The quantitative and statistical analysis of the present study involves the features of the learners' productive prosody previously described in Sections 2.5-2.8. In order to reach the objectives and to answer the main research questions formulated in Section 1.1.1, three steps are involved. In a first step, the productive prosody features are looked at in isolation and then compared to the native control's performance in order to make out deviances and target-norm approximations of individual performances. In a second step, intra- and intergroup variations are investigated. In a third step, all learner varieties are taken together and compared to the two native populations in order to find universal tendencies that all learners share. While some learners from the different L1 backgrounds might perform in a native-like manner for one prosodic feature (e.g. f0 level), they might at the same time perform differently for another feature (e.g. tunes or intonational phrasing), depending on the prosodic features used in their native language. Of course, many other factors might play a role in the preference for certain prosodic features, such as the speaking style, the interviewer, their personality, etc. In the present study, the learner's performances are compared and correlations between all learner variables and recording variables are computed to show significant co-occurrences across all learner groups.

Table 11 summarizes all prosodic features that are analyzed and that represent the dependent variables (or outcome variables) of the present study.

Table 11. Prosodic features, their formulas, units, variable type, and intonational dimension

	Prosodic features	**Definition/formula**	**Unit**	**Variable types**	**Intonational Dimension**
1	Intonational phrasing	IU frequency IU duration (length in seconds) IU speech rate frequency of tones per IU frequency of ips per IP	per hundred words (phw) words per second (WPS)	numeric variables, continuous ratio, interval variables	distributional and realizational + functional (pragmatic functions)
2	Tones Tunes	frequency of tone labels: pitch accents: H*, ^H*, !H*, L*, H+L*, L*+H phrase accents: H-, ^H-, !H- and L- boundary tones: %H, H%, ^H%, !H%, and L% Tone combinations: H* L-L%, H* H- H%, etc.	per hundred words (phw) in % how often are H and L tones and tunes used by corpus (relative frequencies)	categorical variables	systemic and distributional + functional (pragmatic functions)
3	F0 range	Level and span, mean f0, Long-Term Distributional (LTD) measures and linguistic measures (cf. Patterson 2000)	Level: Hz Span: Hz, ST global vs. local f0 range/ LTD vs. linguistic	numeric variables, continuous ratio, interval variables	realizational + functional (pragmatic functions)
4	Uptalk	Uptalk frequency, duration, tune sequences and f0 range: level and span Slope of rise and shape of rise (sustained, rising, or dip)	phw percentages of uptalk instances by all final rises (- and %) and by all ips	numeric variables, continuous ratio, interval variables and categorical variables	distributional and realizational + functional (pragmatic functions)

The prosodic features listed in Table 11 were extracted automatically with the help of *Praat* and R-Scripts from the time aligned *TextGrids* created with *Praat*. The prosodic features were analyzed along Mennen's (2015)[65] four dimensions of the intonation system (systemic, realizational, distributional, and functional), i.e. (1) the inventory of structural elements (pitch accents, phrase accents, and boundary tones); (2) the way these are realized; (3) how they are distributed (frequency of tones, combinations of sequences, and intonational phrasing); and (4) which pragmatic functions each intonation phrase fulfills (seven functions, as predictor variable, not outcome variable). As a first step, the findings will be grouped into the four dimensions of LILt, and then the LILt assumptions will be tested. As Mennen (2015) also points out, it is not always easy to classify intonational deviances into the dimensions of the LILt model and that the dimensions can also interact with each other. The following dependent variables (List 2) will be analyzed in the results section:

List 2. DVs

TONES_PHW	tone frequency phw
HL	High vs. low tone labels
IP_PHW	number of IPs phw
IM_PHW	number of ips phw
CV_TOTAL	number of creaky voice instances
WPS_IP	words per second per IP
WPS_IM	words per second per ip
F0	f0 by tone label
F0_LEVEL_MEAN	f0 level by ip
F0_SPAN_ST	f0 span by ip
DID_SCORE	difference in distance score
UPTALK	uptalk vs. non-uptalk
F0_MIN_ONSET	f0 min at rise onset of uptalking tunes
F0_MIN_TUNE_LOW	f0 at tune low of uptalking tunes
F0_MAX	f0 at rise peak of uptalking tunes
F0_LEVEL_MEAN	f0 mean of uptalking tunes
F0_SPAN_RISE	rise span in STs of uptalking tunes
RELATIVE_SLOPE UPTALK	f0 span/duration of rise in seconds of uptalking tunes relative to the entire IP
RISE_DURATION	rise duration in seconds of uptalking tunes

As Table 11 shows, many measurements are given in per hundred words (phw), this normalization procedure was necessary in order to compare the texts of different lengths and IUs (see Section 3.4 for speaker information). The word count is unpruned for all measurements, which means that all words produced by the interviewee include all kinds of hesitation phenomena. How each dependent variable was defined and operationalized will be explained in more detail in the following subsections.

3.7.1.1 Tones and tunes

Following the ToBI annotation system (see Section 3.4 and 3.5), the present study includes *pitch accents*, *phrase accents*, and *boundary tones*, as well as combinations of these. How these prosodic features are defined in the present study is described in the following:

---

[65] She adopted these features from Ladd (1996).

Pitch accent:
- High or low tone or a combination of these two on a stressed syllable occurring anywhere in ips and IPs in between initial and final edge tones. Symbolized by a *.

Phrase accent:
- A high or low tone on an unstressed or stressed syllable at the end or near the end of every ip (technically speaking also a pitch accent if it occurs on the last stressed or second to last stressed syllable). However, if this accent appears near the end of an utterance (no matter if it is stressed or unstressed), it is always symbolized by -.

Boundary tone:
- A simple high or low tone on an unstressed or stressed syllable at the end or near the end of every intonation phrase. However, if this tone appears near the end of an utterance (no matter if it is stressed or unstressed), it is always symbolized by %. Together with the phrase accents, boundary tones are summarized under *edge tones*.

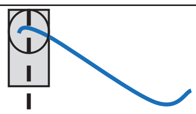
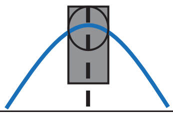
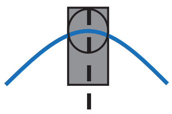
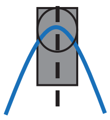
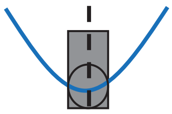
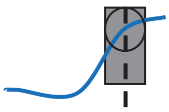
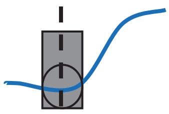
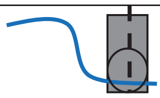
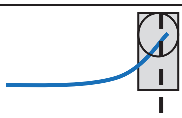
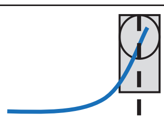
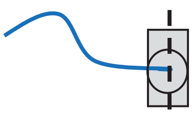
The tone inventories of English (Beckman et al. 2005), Czech (Pešková 2017), German (Baumann et al. 2000), and Castilian Spanish (Estebas-Vilaplana and Prieto 2008, 2010), and their corresponding ToBI labels have been described in Section 2.5.2. Based on a comparison of these four tone inventories and labeling systems, I chose the following combined tone inventory for the present study (see Table 12):

Table 12. Tone inventory used in the present study

Pitch accents	**Phrase accents**	**Boundary tones**
^H*	**^H-**	**^H%**
H*	H-	%H and H%
!H*	!H-	**!H%**
L*	L-	L%
L+H*		
L*+H		
H+L*		

Even though it is a combined tone inventory, the major source is MAE-ToBI, since all recordings are in English. All tones that appear at least in two labeling systems were included in the tone inventory, except H+!H*, which appears both in MAE-ToBI and G-ToBI, since it was rarely used in the recordings of the present study. The only exceptions of tones that appeared only in one ToBI labeling system but that were also encountered in the other varieties of English are ^H% from G-ToBI and !H* from MAE-ToBI. The ^H and !H was allowed in all positions *, -, %. Those that were added for the NN English components (from their respective L1 ToBI systems) are marked in bold font. The final tone inventory thus includes four simple phrase accents, five simple boundary tones, and four simple pitch accents and three bitonal pitch accents (L+H*, L*+H, H+L*), totaling 16 different tone labels. By including the diacritics “!” and “^” for high tones, a more phonetic approach was undertaken for high tones, although such a distinction was not made for low tones. This decision for a more phonetic point of view (at least for high tones) was motivated by the variable nature of L2 prosody and dialectal prosodic variation and to allow comparability across realizational levels (see Orrico et al. 2016 for a similar approach). Stylized versions of this combined tone inventory and the corresponding ToBI labels are represented in Table 13.

Table 13. F0 labels. The light gray cells mean they can be either stressed or unstressed and dark gray ones mean that they have to be on a stressed syllable. The blue line represents the f0 contour. The circle indicates landmark position and the dotted line shows where the point label would be placed in the annotation in *Praat*

No.	Target Label	F0 contour	Description
Initial tones			
1	%H		phrase-initial high boundary tone on accented or unaccented syllable.
Medial tones			
2	H*		local peak on accented syllable in the upper part of the speaker's f0 range for the phrase. This includes tones in the middle of the f0 range, but precludes very low F0 targets.
3	!H*		downstepped high pitch accent on accented syllable, in upper part of speaker’s f0 range for the phrase, lower than the preceding high tone(s).
4	^H*		upstepped high pitch accent realized on accented syllable, in upper part of speaker’s f0 range for the phrase. A very sharp extreme rise, higher than the preceding high tones.
5	L*		local valley on accented syllable in a low or the lowest part of the speaker's pitch range.
6	L+H*		low-high f0 contour. The high tone falls on the accented syllable.
7	L*+H		low-high f0 contour. The low tone falls on the accented syllable.
8	H+L*		high-low f0 contour. The low tone falls on the accented syllable.
Final edge tones			
9	H-, H%		high phrase accent realized on accented or unaccented syllable at ip and IP boundaries.
10	!H-, !H%		downstepped high pitch accent on accented or unaccented syllable at ip and IP boundaries lower than the preceding high tone(s).
11	^H-, ^H%		upstepped high pitch accent realized on accented or unaccented syllable. It is a very sharp extreme rise, higher than the preceding high tones.
12	L-, L%		local valley on accented or unaccented syllables at ip and IP boundaries.

For the combination of these tonal targets, the same rules apply as with the ToBI labeling system described in Section 3.4; thus, each ip has to contain at least one stressed syllable, possibly including a change in f0 level. Each ip has to end with a phrase accent on an accented or unaccented syllable. Each IP simultaneously ends with a phrase accent and boundary tone on an accented or unaccented syllable. That the f0 line does not always appear as straightforward and flawless in the actual recordings can be seen in the *Praat* screenshot in Figure 3. Especially with audio recordings that are not made in a sound-proof room, the f0 line breaks off and does not always reach IU boundaries. The darker shaded areas in the spectrogram represent the stressed syllables, typically including vowels. Every relevant change in the f0 line is documented with tonal targets and set exactly at the highest or lowest point in order to be able to extract the f0 values in Hz for the f0 range calculations.

One of the shortcomings of the ToBI system is that it is not able to represent level tunes. It is not possible to use H*-H-H% for a high level tone and L*L-L% to represent low level tunes. These tunes are used for high rises and low falls, respectively. However, in the present study, these tunes were used to include high rises, high levels, and uptalk (H*H-H%) and low levels, as well as low falls (L*L-L%). That means that sometimes a tune showed the exact same f0 contour from the beginning to the end and was transcribed as L*L-L%, for instance, just as a tune would be transcribed as such which has an overall lowering from the beginning to the end (no matter how small the change). If a clear fall from a higher f0 level was heard and visually verified, a HL-tune was annotated.

The goal is to examine the proportion of high and low tones and tunes in the interactional and monologic discourse. Tones and tunes will be investigated more closely turn-internally and at turn-yielding positions. Intermediate phrases were regarded as turn-internal except when the last ip coincided with an IP boundary. One major drawback in analyzing non-native f0 patterns is that the identification of pitch accents is not as clearly marked as in the NE productions, due to a lower f0 range used throughout utterances (cf. Ramírez Verdugo 2002: 130, 2006a: 18; Ramírez Verdugo and RomeroTrillo 2005: 162).

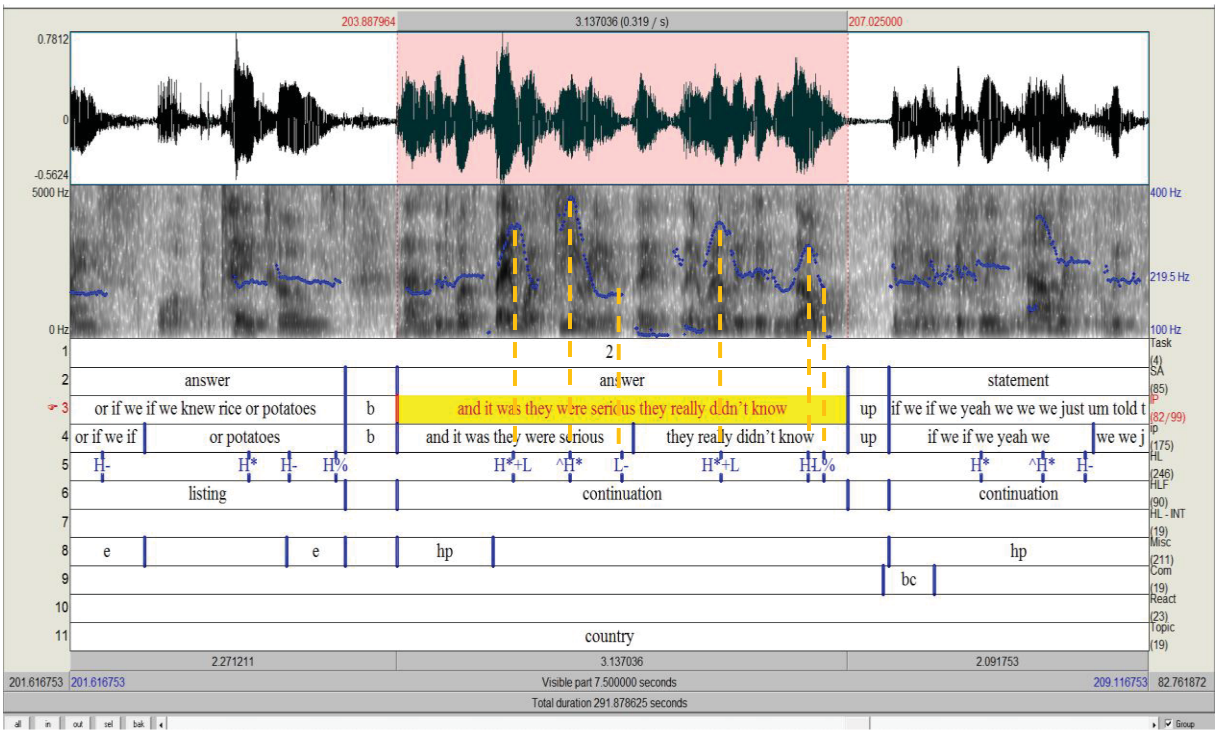


Figure 3. *Praat* screenshot of the annotation of tones (file GE006)

Once the annotation of the ToBI labels was completed, the tone frequencies were extracted automatically with a *Praat* script. Since the native and non-native speakers of English differ in their number of tones and words annotated, the values of the tones are normalized in phw. Relative frequencies in percentages of tones by corpus, discourse function, pragmatic function, age, gender, and other independent variables will be calculated as well. Additionally, the frequencies of tones will be subject to a number of statistical tests described in Section 3.7.3.

3.7.1.2 Intonational phrasing

In the present study, I make use of two levels of phrasing as specified in ToBI, i.e. intonational phrases (IPs) and intermediate phrases (ips). I define IPs as utterances delimited by a particular f0 pattern, i.e. a starting f0 and a final f0, clear f0 resetting, possibly preceded and followed by longer silent pauses. Generally, IPs correspond to sentences in syntactic terms, i.e. to level four break index according to ToBI annotations. IPs can consist of one or more ips. Intermediate phrases are defined as smaller utterance constituents, which generally signal continuation within larger IPs. Intermediate phrases are also delimited by a specific f0 pattern. F0 resetting is generally smaller but still present, possibly preceded and followed by smaller silent pauses. Often ips end with final syllable lengthening - often on filled pauses (i.e. *um*, *er*). Intermediate phrases generally correspond to clauses and phrases in syntactic terms and break index three according to ToBI annotations. How this concept looks in a *Praat* window can be seen in Figure 4. An interval is set for each IP and the corresponding ips. The intervals include the orthographic words uttered by each speaker.

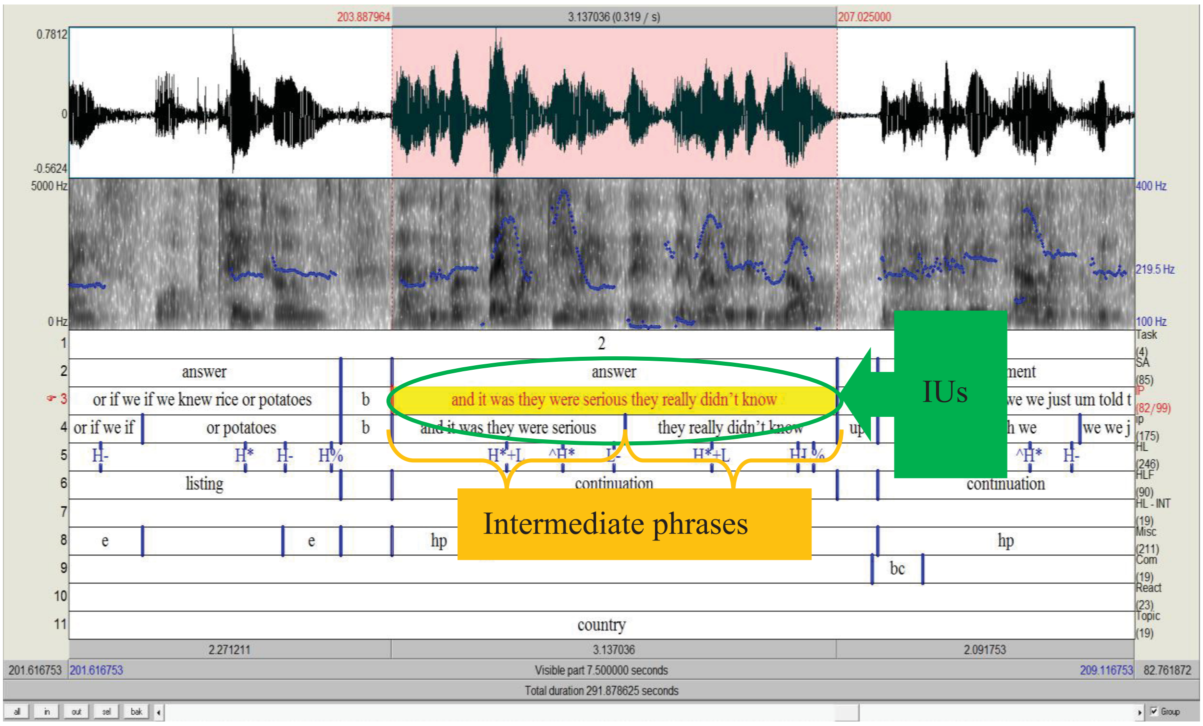


Figure 4. *Praat* screenshot of the annotation of intonation units (file GE006)

The frequency and duration of, and words contained in IUs were extracted automatically with *Praat* scripts.[66] Since native and non-native speech differs in terms of fluency, the frequency of IUs was normalized in phw by the respective speaker sample sizes. The duration of IUs is given in seconds and words. The mean, median, and standard deviations and IQR of words per IU were calculated per speaker. Also, the speech rate was measured by taking the words per seconds (WPS)[67] (e.g. Levelt 1993: 22) within IUs.[68] Thereby I counted contractions as one word, filled pauses also counted as words, and self-corrections were left in as well, i.e. I used the unpruned word counts (see Lennon 1990). Since previous research has shown that while NSs (LOCNE) use repeats less than NNSs in LINDSEI (Götz 2013; Gráf and Huang 2019), however, the study by Gráf and Huang (2019), for instance, showed that 80% of the learners produce repeats in a native-like frequency. Therefore, it should not cause any larger problems to leave repeats in the word count. When WPS was measured by IP, this may include some silent pauses, while most WPS measures per ip would not include many silent pauses, because most ips are delimited by silent pauses. As in the analysis of tones, all independent variables will be checked against the values for intonational phrasing in order to find determining factors for native and non-native prosody. Even though hesitation phenomena, such as filled pauses, repetitions or self-corrections, are not typically in variation analyses, the present study included all of these hesitation phenomena in the total word count

[66] The script was taken from: http://stel.ub.edu/labfon/sites/default/files/point_label_extraction.praat. My special gratitude goes to Wendy Elvira-García for publicly sharing the Praat script for automatic label extraction.
[67] Different speech tempo measures have been proposed, either measures by words, syllables, or phonemes per a certain time unit, i.e. seconds or minutes (cf. Gráf 2019: 176). Even though words per minute is the most frequent measure in learner corpus research (ibid.), none of the IUs in the present study are a minute long and are therefore measured in seconds.
[68] Combining fluency measures with prosodic units is not a novel undertaking; see e.g. Gut 2012 who uses an MLR measured in the mean number of syllables per IP.

of approximately 560 words per speaker. The reason for that is that filled pauses and repeated words, especially in non-native speech, can also be the bearers of tones. The exclusion of these hesitation phenomena from the analysis of f0 patterns would falsify the overall picture of f0 patterns within IUs. Thus, for the calculation of IU duration (in seconds and words) and IU speech rate (in WPS), only unpruned (including hesitation phenomena, cf. Lennon 1990: 406) measures will be considered.

3.7.1.3 F0 range

In the present study, I define f0 range as the variation in f0 values throughout utterances corresponding to speaker-specific minimum and maximum f0 limits. Following Ladd (1996, 2008), I consider f0 range a two-dimensional construct; consisting of an f0 level (overall f0 height) and f0 span (range or extent of f0) (see Figure 5 for a visualization of this concept).

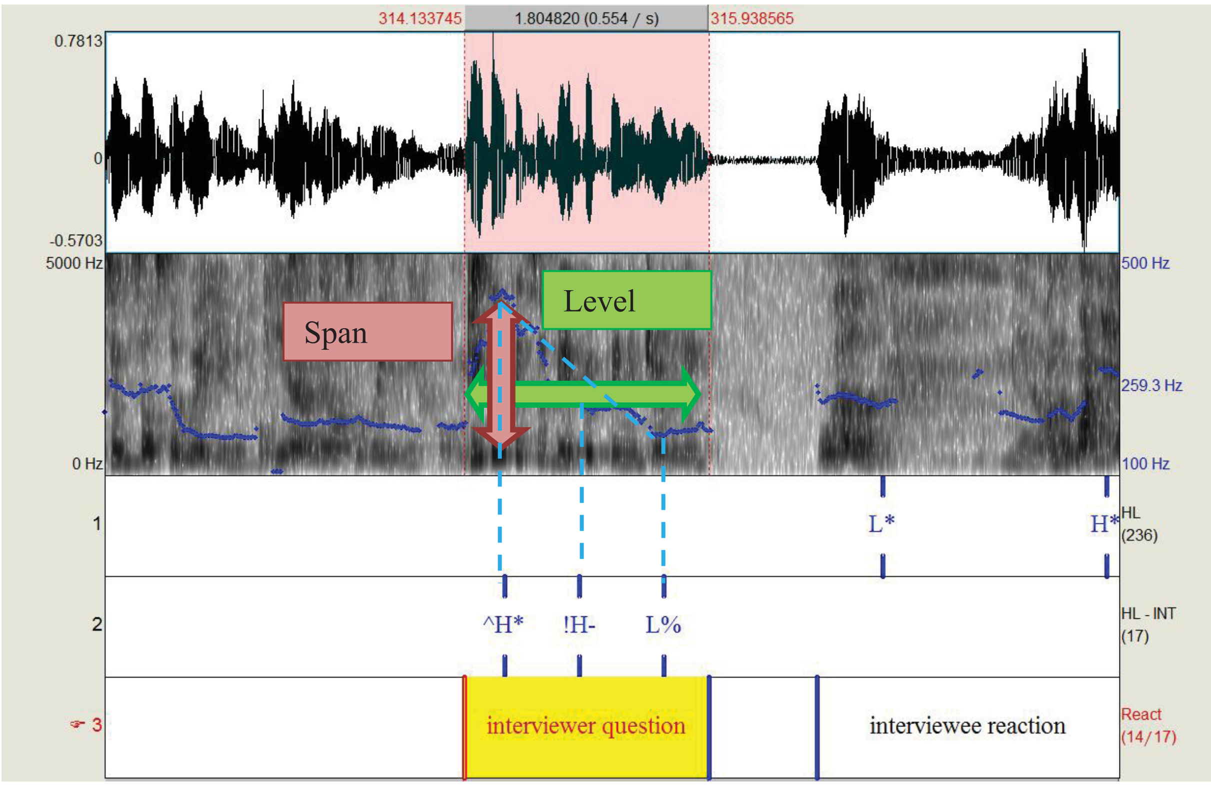


Figure 5. *Praat* screenshot of linguistic measures (f0 level and span) from first pitch accent to phrase accent to the final boundary tone of the American English interviewer with reduced tiers for visualization purposes (file GE029)

During the annotation process, f0 maxima and minima were identified and labeled in *Praat* with the corresponding ToBI labels from the combined tone inventory (see Section 3.7.1.1). Care was taken to place the point label on areas where a clear f0 line was visible and avoiding microprosodic perturbations due to creakiness in a speaker's voice or other octave errors (cf. Beckman and Venditti 2011: 494). With the help of the speech analysis software *Praat* (Version 6.0.43), all f0 measures per tone were extracted automatically with the help of *Praat* scripts.[69] The *Praat* settings were set to a pitch floor of 50 Hz and a pitch ceiling of 600 Hz, and time steps of 10 ms (*Praat*'s standard setting for time steps), for both male and female speech due to inter-gender variation within all corpora. All extracted f0 contours and their corresponding Hz values were visually inspected and manually corrected. Even though all tone labels were included in the tones and tunes analysis, for the f0 range analysis data-trimming was necessary. For instance, cases where the f0 line was unrecognizably masked with background noises or overlapping speech had to be excluded. Also, f0 contours in *Praat* in creaky voices are generally considered unreliable (cf. Beckman and Venditti 2011: 493) and so these f0 values of 75 Hz and below resulting from creaky voice where deleted. This step was necessary to avoid a strong effect on the overall f0 range measurements with extreme low f0 values. Thus, even though a low tone was auditorily perceived and visually depicted in the spectrograph, a low tone was only annotated in *Praat* but the f0 value was nevertheless excluded from the analysis. These f0 measures were extracted for every analyzable IP in the corpus, as well as every uptalk instance (see Section 3.7.1.4). By analyzable IPs I refer to f0 patterns that have to be present and cannot overlap with interviewer speech or other background noise in critical areas. As a first step, all f0 measures of all linguistic labels were extracted automatically with *Praat* scripts and then discarded if the criteria for analysis were not met, e.g. if f0 values of 50 Hz are extracted or if an interviewer interruption was annotated on a parallel tier. In this fashion, each annotated speech sample was examined for inclusion in the f0 range analysis. All outliers of unusually high and low f0 values, i.e. under 75 or over 350 Hz, as well as spans over 200 Hz and negative spans, were inspected and corrected if necessary. In addition, any low tones above 250 Hz and high tones below 100 were inspected and corrected manually. In order to get complete f0 range measures by IUs, if one of the f0 measures returned from the *Praat* scripts was 50-75 Hz for one IP and/or rendered unreliable do to creaky voice, the entire IP was deleted. This data trimming procedure resulted in the deletion of 134 IPs of 6,624 IPs from the entire corpus (1.86%). However, these IPs were not removed from the tone label analysis. The remaining 6,490 IPs form the basis of the f0 range analysis. Additionally, on a speaker-by-speaker basis, data points with residuals deviating by more than two standard deviations were excluded from the analysis, if a normalized distribution is needed for the respective statistical model used.

After the removal of outliers, as described above, global and position-sensitive f0 range measures by ip were employed:

List 3. List of tunes and their ToBI labels

- Flat tunes: All H-H and L-L were left out (because that indicates a relatively flat intonation. However, this was left in the global f0 range measures to see if non-native speech is flatter overall).
- High low vs. low high (non-final ips): H*L- and L*H- (earlier in an IP)
- High low vs. low high (final ips): H* L-L% and L*L-H%/ L* H-H% (later in an IP)
- High-low-high vs. low-high-low (non-final ips): H* L*H- and L* H*L-
- High-low-high vs. low-high-low (final ips): H* L-H% and L* H-L%

[69] The script was taken from: http://stel.ub.edu/labfon/sites/default/files/extracts_f0_from_points.praat. Again, my special gratitude goes to Wendy Elvira-García for publicly sharing the Praat script for automatic label extraction.

Only one minute of speech has been shown to be sufficient to determine a speaker's f0 range, since the standard deviation converges fairly quickly to within two Hz if the mean is compared with the median f0 (e.g. Horri 1975: 195; see also Nolan 1983: 123). Since all speakers within the present corpus reached at least one minute of speech or more, the f0 range measures will be representative of their overall f0 range.

In order to find differences and similarities between native and non-native speech, speaking styles (dialogues and monologues), and discourse functions (statements and answers), various measurements are taken from the present data set. The more concrete and fine-grained the prosodic analysis, the greater will the tendency be for differences between native and non-native speech to emerge. Therefore, f0 range was operationalized as long term distributional (LTD) measures based on the distribution of f0 within a given speech sample and linguistic measures, i.e. f0 level (mean, median) and f0 span (min, max, SD). A summary of all f0 range measures is presented in Table 14.

Table 14. F0 range measures (cf. Mennen et al. 2012: 2254)

	LTD	**Linguistic**
Level	mean f0 Hz	By tone label analysis (conflated):
	median f0 Hz	%H, H*, H-, H%, L*, L-, L%
	f0 maximum Hz	By tune analysis in final and non-final position:
	f0 minimum Hz	HH, LL, LH, HL, HLH, LHL
Span	SD/IQR	By tone label analysis (conflated):
	max-min ST	%H, H*, H-, H%, L*, L-, L%
	(PDQ)	

These measures in Table 14 were taken in order to ensure comparability with results of previous f0 range studies (e.g. Mennen et al. 2012, 2014). The f0 level is typically measured as the mean and median in Hz. Even though standard deviations will be taken, it has to be kept in mind that SD measures were primarily designed for symmetrical data. However, f0 values are often asymmetrical, i.e. skewed to the right (e.g. Mennen et al. 2012: 2255). Following, Patterson's (2000: 123) recommendations, the f0 span will be measured both in Hz and STs. Nolan (2003) uses a psycho-acoustic span-matching task to assess different scales to represent the relative perceived differences between male and female speech, he finds that STs and ERBs performed best, with STs performing marginally better. Therefore, f0 span will be analyzed in STs. F0 measurements were transformed from Hz to ST with the formula 12*log2(Hz/HzMin), where HzMin is the minimum f0 of a speaker in an ip. Previous research has found that male and female speakers exhibit differences in f0, with female speakers having an f0 that is between 75-90% higher than the f0 of males (Hollien and Paul 1969; Hollien and Shipp 1972, as summarized in Diehl et al. 1996). F0 of female speakers was 54-83% higher than the f0 of males in the present study. Keeping this in mind, the f0 range differences in male and female speakers were addressed by treating them as separate groups of speakers and separating them for each further statistical analysis. The quite large inter-gender f0 variability for some corpora did not allow for an approach of normalizing f0 by gender, as it is done in some f0 range studies (e.g. Fuchs 2018).[70] In general, the NSs had an increase around 70% from male to female speech in f0 level median (LOCNEC 78% and AmE corpus 72%). The CzE and MurE speakers had a smaller increase of around 50% (MurE 50% and CzE 54%). The GerE and MadE speakers have the largest disparity of around 80% (GerE 83%, and MadE 88%). Therefore, I decided to only include ST measures to normalize the data, since this measure (and ERB conversion (Moore and Glasberg 1983)) is used to level vocal tract differences between male and female speakers (e.g. Henton 1989; Armstrong et al. 2016).

All possible tunes (see List 1) are computed for global as well as local f0 range. A global measure would be for instance from the first peak (either %H or (^)H*) to the final low L%. A local measure would be from the first peak (either %H or (^)H*) to the next phrase accent (either !H- or L-) and from the phrase accent to the final low boundary tone (!H - L%). See Figure 5 for an example. However, many IPs in the present data set do not exhibit such a perfect declination f0 contour, and sometimes more than one peak can be found in ip-medial position. In most cases there would be one IP-initial high and one final low. However, there can be various numbers of peaks and/or valleys in between, depending on the length of the IPs. For instance, many learner f0 patterns can look like this for one intonation phrase: H* H* ^H- H* H* L- H%. Therefore, a by ip analysis will be undertaken, i.e. H* H* ^H- (conflated to HH-tune) measures and H* H* L- H% (conflated to HLH-tune) measures, just to show an example from the longer IP.

F0 range has been shown in the theory to be dependent on various aspects: speaking style, discourse, topic shifts, location (local vs. global), and entrainment (interviewer influence). In order to analyze prosodic entrainment, i.e. whether speakers match their f0 range (level and span) at speaker transition points after taking over the floor, the following speaker combinations are considered as well:

- L1 (A) - L1 (B): native interviewer and native interviewee (e.g. NSV)
- L2 (A) - L2 (B): NN interviewer and NN interviewee (e.g. LINDSEI-CZ)
- L2 (A) - L1 (B): NN interviewer and native interviewee (e.g. LOCNEC)
- L1 (A) - L2 (B): native interviewer and NN interviewee (e.g. LINDSEI-GE, CZ, SP)

Based on results from previous research on prosodic entrainment, matching of turns should appear early on. Therefore, the last ip of speaker A (interviewer) and first ip by speaker B (interviewee) is compared directly in f0 range measures. Also, the entire IPs are compared in terms of f0 range and, therefore, global as well as local entrainment can be analyzed. Therefore, the hypothesis is that interviewee f0 range will be wider in the immediate answers to interviewer questions or statements (post-interaction), than in parts which are not directly preceded by interviewer stretches (pre-interaction). To calculate the amount and direction of accommodation, a difference in distance (DID-score) measurement was calculated (cf. Babel and Bulatov 2012). First, the distance between the interviewer and interviewee's f0 values for

[70] Where the mean and median f0 values of the female speakers are divided by the mean of an increase by the respective percentages for each corpus (cf. Fuchs 2018 for a similar approach, he calculated an increase of 1.82 for his ESL data).

each ip was calculated. Second, the f0 values from the baseline pre-interactions were subtracted from the post-interaction f0 values.

> A negative difference in distance value indicates that the difference in f0 between a participant and the model talker shrank as a result of f0 accommodation. A positive difference in distance value indicates that the f0 difference increased, which would signal divergence with respect to f0 […] [A] value of 0 would indicate no change in participants' f0 […]. (Babel and Bulatov 2012: 238-239)

For each speaker a mean and median value for all measurements was calculated. However, for speakers who had less than three tokens whenever the data would be subsetted, e.g. only answers given vs. statements given, and if there was only one answer for one speaker and 25 statements, then this speaker would be dropped entirely from the analysis, since there is not a reliable comparison to be made. The difference in distance measure and all other f0 measurements mentioned in this section were used as dependent variables in a mixed effects model where speaker gender (male or female), interviewer gender (male or female), speaking style (dialogue or monologue), interaction (pre- and post-interaction), and region (AmEO, CzE, GerE, MadE, MurE, NBrE, and SAmE) were the independent variables.

In addition to the entrained passages, the f0 measures of quoted speech (also known as direct reported speech) samples have to be accounted for as well. According to Bolinger (1946: 197), when speakers use quoted speech, their intonation tries to "re-enact the original intonation". Previous research found that quoted speech has an overall greater f0 range than other parts in a narrative (e.g. Klewitz and Couper-Kuhlen 1999; Wennerstrom 2001; Tyler 2014). Tyler's (2014) study with the Santa Barbara Corpus of American English has shown that quoted speech starts and ends higher, rises more overall, has steeper slopes but is no different in duration. He found quoted speech to be gender-dependent. Jansen et al. (2001) found in an analysis of the Switchboard corpus that directly reported speech is characterized by a greater overall f0 range than the surrounding narrative material and that it is preceded by IP boundaries. However, they also found that indirectly reported speech does not distinguish itself prosodically from the surrounding narrative. Klewitz and Couper-Kuhlen (1999: 468) also claim that "which specific prosodic format is used ultimately depends on the circumstances of the context in which the reported speech is situated". Marking certain ips or parts of ips as reported speech or imitations of other speakers' voices was undertaken in the present study to find out more differences about prosodic structure and function in native and non-native speech. How this was done is described in the manual in Appendix 1. Generally, the following method was applied while annotating each recording: "Where a prosodic shift begins, quoted speech begins; where this shifts ends, the quoted speech ends" (Klewitz and Couper-Kuhlen 1999: 469). However, where reported speech was marked verbally only, without prosodic boundaries, these passages were also marked as "imit", standing for imitations, different voices, and/or reported speech (including both direct and in-direct speech). Regardless, of prosodic boundaries being present or not, all quoted speech passages were annotated. However, a distinction between these different types of quoted speech was not done. See Figure 6 for an example of quoted speech in the present data set.

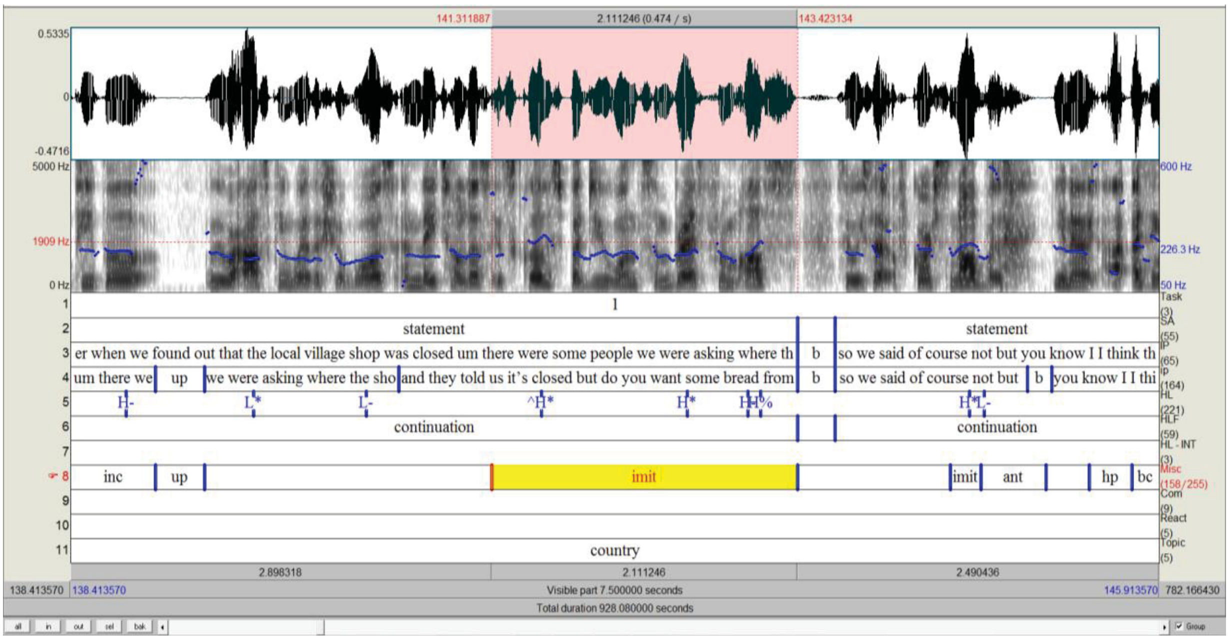


Figure 6. *Praat* screenshot of a quoted speech sample (file CZ001)

As one can see from the screenshot above, a by ip analysis of quoted speech is not possible and, therefore, only a by tone label analysis was performed, because very often only part of an ip was marked as a quoted speech sample. As in the first ip that starts with "and they told us", which is spoken in the speakers usual f0, and the second part of the ip is produced with a higher f0 "it's closed but do you want some bread from us" and then in the next ip the speaker produces another quoted speech sample "so we said" is in the speakers normal f0 and "of course not" in an imitated speech, and "but" is again in the normal f0 level. All questions or question-like ips were excluded from annotation, the only possible questions stem from quoted speech such as in the example above "do you want some bread from us?". Quoted questions, were, however, very rare and were not distinguished during the intonation process from non-question quotations.

3.7.1.4 Uptalk

Uptalk information was extracted manually, because of many of the backchanneling overlaps with all tunes, which called for manual checking of each tune. The presence or absence of uptalk in the present sample was based on the definition of uptalk by Warren (2016: 2, emphasis added): "a *marked* rising intonation pattern found at the ends of intonation units realized on declarative utterances, and which serves primarily to check comprehension or to seek feedback". Based on that and in the context of the present study, I defined uptalk as follows: a rising intonation in utterance-final position within declaratives, eliciting some type of verbal or non-verbal minimal-response from an interlocutor, which can appear especially in confirmation requests/checking and floor-holding/continuation. As described in Section 3.5, all type of questions, and tag questions were removed from the annotation as well as the analysis. Small one-word phrases such as *yes*, *okay*, *no*, were excluded from the analysis of uptalk, as were incomplete phrases. List intonation was included in the analysis of tonal patterns. However, there are only very few items that were marked as uptalk within listing intonation in this sample. List intonation appears standardly with rising intonation in SBrE (cf. Levon 2018: 5). The uptalk tunes were only annotated within the 560 word samples from each speaker, and no additional tunes were included from beyond this 560 word sample. Thus, either more uptalk or no uptalk features could possibly be found and analyzed in the corpora of the present study.

The uptalk instances in the six corpora were analyzed as follows. In order to describe the form of uptalk, the AM descriptions of tunes are used. The following pitch accents and edge tones were identified in the uptalk tunes:

Form and distribution:

- pitch accents: H*, L*, !H*, ^H*, H+L* and L*+H
- edge tones: L-, H-, !H-, ^H-, and L%, H%, !H, ^H%

In the present study, I allowed both H* H-H% and L* H-H% as types of uptalk tunes. IPs labeled L* H-^H%, L+H* H-^H% or H* H-^H% were classified as uptalk instances, in accordance with a suggestion made by Ladd (1996). In order to assign these AM labels to the uptalk tunes, I made use of *Praat*'s f0 tracking and auditory perception. As the variety of research on Australian uptalk tunes suggests, it is necessary to focus on all kinds of tunes in uptalk: simple, complex, and compound (cf. Fletcher and Loakes 2006: 42). For this reason, my analysis also includes single uptalk tones on one accented syllables and complete tunes of up to three single and bitonal tunes. Thus, simple high rises would for instance be H-H%, and L-H% would be an example of an expanded-range rise, because the starting point is lower than H-H% and the range is expanded (cf. Flechter and Loakes 2006: 42-43 for these terms). The frequency of these patterns will be determined. In order to find differences in the realizational dimension between the native and non-native varieties of English, the following additional measures were taken:

Realization - phonetic details (based mostly on Armstrong et al. 2016 and Levon 2018):

1. Duration of rise (in ms) - rise onset (lowest point) in the tune and end of rise (highest point, i.e. its peak)
2. Duration of entire uptalk tune
3. Starting vs. ending pitch
   - F0 at onset of rise (f0 minimum)
   - F0 at peak of rise (f0 maximum)
4. Pitch excursion (difference between f0 max. and min. in ERB) in absolute and relative terms (Di Gioacchino and Crook Jessop 2011)
5. Slope of the rise (dividing the total pitch excursion by the duration of the rise (onset-peak).
6. Pitch dynamism quotient (pdq; how fast pitch changes, e.g. "pitch dynamism" Henton 1995). The pdq, is calculated by taking the SD of the f0 values of a speaker and dividing it by mean f0 level in Hz (cf. Hincks 2004; SD/mean). The pdq measures a speakers' variation in their pitch over the course of an utterance. It is considered a measure of "liveliness" (Hincks 2004).
7. Shape of the rise (sustained, dip, or rising): What I call "dip" here is also termed a "split fall-rise contour" by Fletcher (2005), e.g. ^H* H- or ^H* L* H-.

8. Duration and type of pauses following uptalk tune

Examples of uptalk tunes can be seen in Figures 7-10.

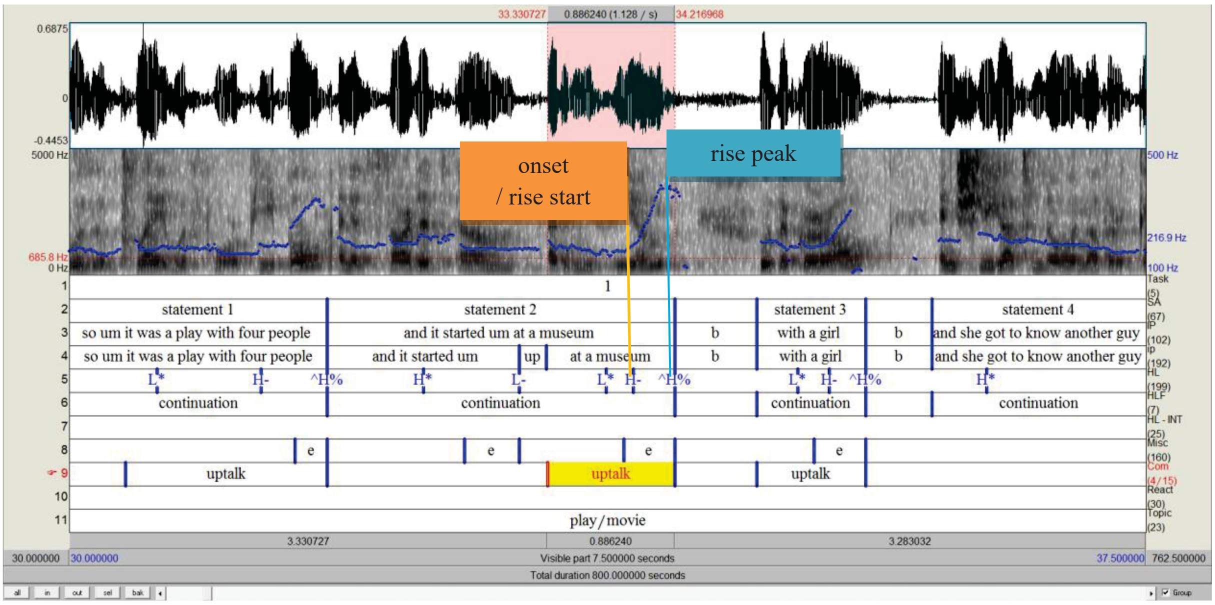


Figure 7. *Praat* screenshot of uptalk example with onset and rise peak (file GE032)

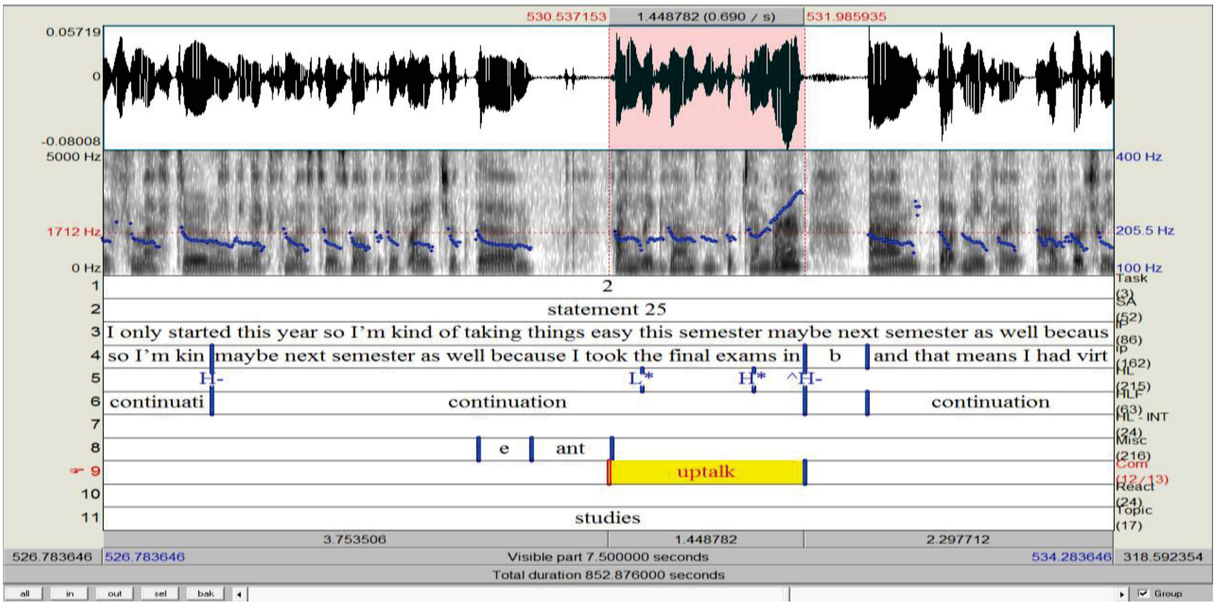


Figure 8. *Praat* screenshot of an entire uptalk tune with a sharp rising slope - selected area is the length (1.4 seconds) of the entire uptalk tune that was measured (file CZ021)

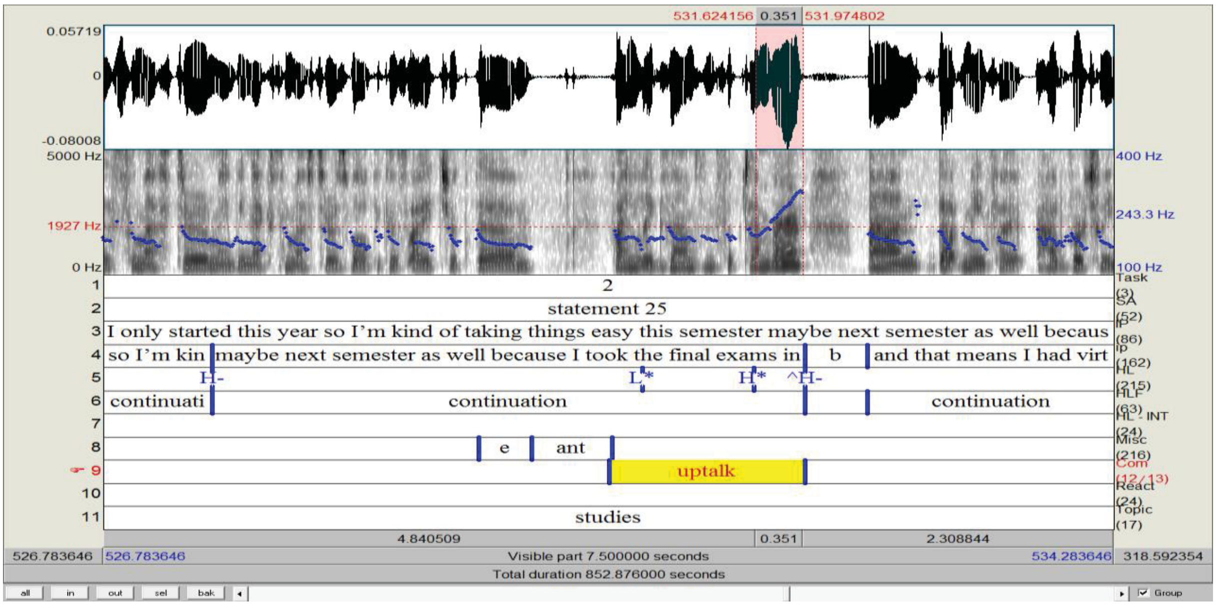


Figure 9. *Praat* screenshot of how the length (0.35 seconds) of the rise was measured, which is part of a larger tune (file CZ021)

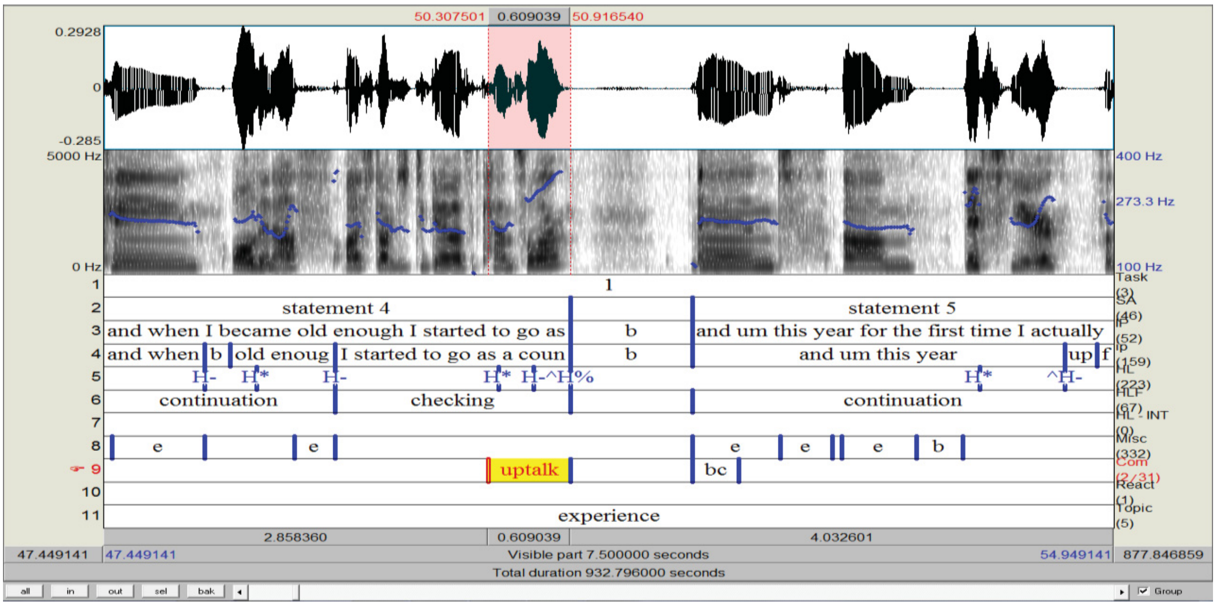


Figure 10. *Praat* screenshot of an extreme uptalk rise, followed by a long breathing pause and backchanneling by the interviewer, and a non-uptalking rise in the following ip (file CZ032)

As can be seen in Figure 7 the rise excursion is measured as the difference between the f0 at the peak (i.e. highest reliable point of the rise) and the f0 at the onset of the rise (starting point of the rise), also called "elbow" of the rise which shows the start of a clear upward inflection of the tune (cf. Arvaniti and Ladd 2009: 55). This is an analysis of the f0 span within the rise itself; however, I also included the overall f0 range of the entire phrase as well. The highest point remained the same but the lowest point in the overall tune was chosen as well. Fletcher (2005: 1381) also argued that not only terminal elements of uptalk tunes need to be considered, but the entire tune of the intonational constituents needs to be investigated as well. Many intonation researchers do not take this into account and only investigate terminal constituents of an uptalk tune. According to Di Gioacchino and Crook Jessop (2011), the

ToBI labeling system is not equipped to appropriately describe uptalk tunes and their phonetic variation. In the ToBI annotation system, all tunes ending in L-H% all rising end-of the sentence tunes get lumped together, including questions, continuation rises, etc. They suggest that a more accurate description of uptalk could be achieved by using relative excursion sizes, which includes an analysis of the speaker's f0 span in relation to the speaker's overall f0 range of the immediately surrounding IPs: "It is possible that it is not the absolute value of the excursion, which we perceive, but instead the height of the excursion in proportion to the speaker's overall pitch range" (Di Gioacchino and Crook Jessop 2011: 2). This relative excursion size is obtained by calculating the absolute f0 span of the rise (f0 max-f0 min) in proportion of the overall f0 span of the prosodic environment (i.e. the IP in which it occurs) (ibid.). This is achieved by dividing the rise f0 span by the total IP f0 span. For instance, if we have an absolute rise span of 13 STs, which is contained within an IP of an overall f0 span of 2 STs, then it would be assigned a relative f0 span of 6.5 STs.

These phonetic features described above will be combined with an analysis of the phonological representation (ToBI labels) of the uptalk tunes, as well as a function analysis of the pragmatic functions. By doing so, all four intonational dimensions proposed by Ladd (1996, 2008) and Mennen (2015) can be investigated in native and non-native speech. In addition to that, the context in which the uptalk tunes occurs will also be considered. Since previous research (Levon 2016) has shown that uptalk is placed differently by men and women in narratives, I will also take into consideration the position in which the uptalk tunes occur in the monologues, i.e. are produced at the beginning or end of the selected samples. In the present study, I did not annotate the entire narratives. Sometimes a 560 word sample might include only the beginning of a narrative and sometimes a complete narrative. However, I did include the information in the metadata spreadsheet that if a narrative was annotated from the beginning or later in the recording. This will help to identify in which positions in a narrative or dialogue uptalk is more frequent. Levon (2016: 146) in his pragmatic analysis also included the position of uptalk in a turn. I followed this procedure by annotating whether the uptalk tune is immediately followed by an interviewer question or whether the floor is held by the interviewee. This will show that most occurrences of uptalk are not interpreted as questions or requests of any sort, and that the listener does not react to the tune by answering but rather by minimal responses such as backchanneling, which do not threaten the speaker's turn.

Additionally, every uptalk token was classified for its discourse and pragmatic function and coded for two discourse functions (statements and answers) and four pragmatic functions (continuation (also referred to as floor hold), listing, checking (also referred to as confirmation requests), insecurity (to express reservation)). Even though at times more than one pragmatic function can apply to each ip, I chose the most salient one for each ip based on the context. For instance, if a speaker lists items, such as countries he or she has been to, it could also be annotated as a continuative phrase, but these were consistently classified as listing phrases. Problems with the annotation of uptalk functions have been reported by Park (2011), who had three NSs of English annotate NNSs' recordings for accuracy of intonation. The native annotators were unsure whether the NNSs were using uptalk or signaling continuation (Park 2011: 74-75). See Colantoni et al. (2016: 15) for a similar problem with determining the pragmatic meaning of utterances in spontaneous speech. According to Warren (2016: 189), "we can see uptalk as an extension of the existing intonational system.

Critics, however, often fail to do this, and treat uptalk as though it were fulfilling the functions already attributed to rises" (Warren 2016: 190) and he continues with the following: "continuation rises and lists should not be confused with uptalk, they have to be distinguished from uptalk on both formal and distributional grounds". This, however, is not a straightforward task since many researchers acknowledge the fact that uptalk is also used for floor holds and continuation purposes (cf. Arvaniti and Atkins 2016) and the uncertainty of the native labelers from Park's (2011) study further confirms this issue.

How these discourse and pragmatic functions were determined is explained in Section 3.7.2. However, in the context of uptalk, the function of "continuation" has to be explained, since continuation rises are to be differentiated from uptalk rises. In the present study, continuation rises were non-dramatic rises at the end of IUs which usually did not try to elicit any type of feedback from the interviewer. Therefore, the pragmatic function of "continuation" could also be called floor-hold as in previous uptalk studies. However, it is not just a floor-hold but at the same time also a confirmation request, i.e. a request for mutual comprehension and opening up the possibility for clarification requests, etc. Thus, as one can see, determining the pragmatic functions for uptalk tunes was not a simple matter. In order to make the present study comparable to previous and recent uptalk research (e.g. Ritchart and Arvaniti 2014; Arvaniti and Atkins 2016), uptalk tunes will be investigated within their discourse and pragmatic functions within ips.

Other variables will be investigated as well in order to determine whether these have an influence on the form and function of uptalk tunes, i.e. the influence of the interviewer (questions asked right before and after uptalk tune) and to see whether the interviewer might have interpreted the rising intonation as a question instead of a statement. Also, whether uptalk elicits more backchanneling from the interviewers than other types of high tunes will be considered. The independent variables mentioned in Section 3.7.2 will also be included in the analysis of uptalk tunes of native and non-native speech.

3.7.2 The independent variables (predictor variables)

As the theory sections have shown, prosody is highly variable and it is quite difficult to determine the contexts in which particular tones and tunes could be predicted to occur. However, it is rare that other studies include any other variables or contexts that might explain their results. The possible combinations and interactions between prosody, their contexts, and functions are almost infinite (cf. Jilka 2007: 87). Therefore, the present study attempts to account for this high variability and analyzes the productive prosody features displayed in Table 11 in Section 3.7.1 in order to find out which factors are responsible for the differentiation between native and non-native speech. The prosodic features presented in Table 11 in Section 3.7.1 were then analyzed concerning the following independent speaker variables across all seven regional varieties of English:

1. L1: NBrE, AmE (SAmE and AmEO), CzE, GerE, SpE (MadE and MurE)
2. Nativeness: native vs. non-native
3. Sex: male or female
4. Age (numeric variable)

In a quantitative approach, the study sets out to explain the variation in f0 and intonational phrasing by including sex, situational, and contextual factors. The following interview variables are considered in the analysis (all are categorical variables):

1. Speaking style:	mono, dia, both
2. Later or earlier part annotated:	beginning or later
3. Emotional speech:	emotional, factual, both
4. Good experience or bad:	good, bad, both
5. Gender of interviewer:	male or female
6. Interviewer intonation	number of tonal landmarks and tunes
7. Familiarity with interviewer:[71]	unfamiliar, vaguely familiar, familiar
8. Topic:	country, experience, entertainment, self
9. Topic initiations:	numeric variable
10. Interviewer influence	tones, length and type of stretches (questions/comments)

The predictors, topic, speaking style, and emotional speech specifically need further explanation at this point. The topic choice will be analyzed more closely in the quantitative part. Each topic is then compared in terms of f0 range. Due to data sparcity issues, a detailed by-topic analysis (which was initially undertaken at the annotation phase, see Appendix 1 for the list of detailed topics) could not be maintained and most topics were conflated to four general categories: The topic country includes (country and country 2), experience (experience from LINDSEI/LOCNEC, and childhood experience from NSV), entertainment (books, plays, and movies), and self (studies, food, political opinions, jobs, future plans, family/friends, leisure, and self).

Additionally, there were three different tasks possible for each LINDSEI and LOCNEC recording, i.e. task (1) the monologue, (2) the informal interview, and (3) the picture description. For the selected stretches in this study, only tasks 1 or 2 were chosen. Topic shifts are most often initiated by the interviewer but can also be initiated by the interviewee him/herself. The beginning of new topics, just as the initial reactions to interviewer questions and comments, is expected to have a wider f0 range than continued topics. While there are participants who only speak about a country for the entirety of the annotated speech sample (typically the uninterrupted monologues), there are some participants who constantly switch from completely different topics (typically in the dialogues). Thus, in a first step, the data will be analyzed according to the broader topic distinction of country, movie, and experience and then the prosody production of those with one continuous topic vs. those with many different topics will be compared. Thus, an annotated stretch can include both task 1 or 2, or just one of the options. Stretches that were annotated only within task 1 are mostly monologues but can also include interruptions or questions from the interviewer and those stretches annotated within task 2 can also include monologic speech (i.e. longer answers to questions) but are mostly dialogues. The topic was determined from the general classification as explained in the LINDSEI manual, which can be about a country, movie, or experience. Nevertheless, the recordings can be classified into

[71] Available for all corpora in the data set except LINDSEI-GE, NSV, and NWSP.

more detailed distinctions of topics and topic-shifts, which might have an influence on the production of certain prosodic features.

The speaking styles were determined by the involvement of the interviewer. Some annotated stretches include pure monologic speech without any interruptions by the interviewer. Then there are some that have a few interventions from the interviewer but still continue as monologues. These were also classified as monologues. Stretches that are clearly in a longer question and answer form were annotated as dialogues; however, at times they can also include longer monologic stretches. In a few cases, there are stretches that include an equal distribution of monologic and dialogic speech and were thus annotated as "both". Then, additionally, it was noted whether the selected stretch was produced at the beginning of the recording or annotated after a few minutes. The fluency of the speakers decided on which part of the recording would be annotated. Additional notes were made about whether the speech was emotional or not (yes or no) and whether the participants speak about a rather good or bad experience (yes or no). Nevertheless, in Swerts and Hirschberg's (2010: 1338) words "it is hard to establish whether the annotated emotions were indeed those experienced by the original speakers" and are therefore subjective.

These speaker and interview variables were compared within the first type of CIA analyses (NS vs. NNS). The following learner variables were selected to compare the CzE, GerE, and SpE speakers within the second type of CIA approaches (NNS vs. NNS):

1.	Age	(numeric variable)
2.	Years of English at school	(numeric variable)
3.	Years of English at university	(numeric variable)
4.	Months in English-speaking country	(numeric variable)
5.	Stay abroad/country:	short stay, long stay, no stay abroad/USA, Australia, GB, Wales, Scotland
6.	Other foreign languages	(numeric variable)

Most of this information was extracted from the LINDSEI, LOCNEC, and NSV/NWSP databases, except information on the speaking style, stretch in an earlier or later part of the recording, emotional speech, description of a good or bad experience - all of which were elicited during annotation. Since I made use of existing corpora, I had no control over the balancing of gender, age, topic, stays abroad, etc. There are clearly more female speakers in all corpora except the NSV, which is the only balanced corpus in terms of gender and topic. The male-to-female ratio in the different corpora used is 1:4 in the three LINDSEI corpora, 2:3 in LOCNEC, and 1:1 in the AmE corpus (consisting of NVS and NWSP recordings). However, whenever possible a balance was achieved by choosing similar stretches of annotation, e.g. beginning or later.

Additionally, paralinguistic features that were annotated on the miscellaneous tier (Misc Tier), which all qualify as categorical variables, can be used as independent speaker variables as well:

1.	ant	anthrophonics, e.g. cough, speech-laughs and non-speech laughs, etc.
2.	bc	backchanneling
3.	cv	creaky voice

4. e elongation
5. fr foreign words
6. hp hesitations including repairs, repetitions, etc.
7. imit imitations of others, accent modification
8. int speech by the interviewer
9. intr interruptions
10. whis whispered speech

The frequency of these paralinguistic features is also compared and possible correlations are determined. In the present study the prosody of laughs will not be investigated by itself (see Trouvain 2014 for such an approach). However, in the present data set, utterances were sometimes made that included speech laughs (when laughter is produced simultaneously with articulated speech: Trouvain 2001). Non-speech laughs were not analyzed.

In addition to these various independent variables, a functional dimension was also included by differentiation between different discourse and pragmatic functions of each ip. The binary differentiation of the discourse functions of utterances into “statement” and “answer” speech acts by the interviewees will further help to determine prosodic patterns and prosodic entrainment. The overall question is whether the f0 range is wider after interviewer questions or comments. Thus, the interviewer stretches are analyzed in their f0 range side by side with immediately following interviewer stretches. It will be checked whether the neighboring IPs approximate to each other in their f0 range.

Since previous studies have shown that assuming a standard intonation pattern for each sentence type based on literature reviews and models of intonation does not depict a correct picture of the production of prosody (cf. Puga 2018; see the discussion in Section 1.5.4), the present study takes the NBrE NS mean as the main comparative value for the learners’ performance, although the AmE means will be used as controls. Different statistical methods (described in the following section) will reveal any kind of patterns in all native and non-native prosodic and non-linguistic variables. In the statistical models I will use the following abbreviations for the various IV (see List 4):

List 4. IVs (sociolinguistic, contextual, and linguistic)

socio-linguistic variables	AGE	speaker age (interviewee)	numeric
	REGION	variety of English	7 levels
	SEX	speaker sex (interviewee)	male vs. female
learner variables	NO_OFL	number of foreign languages	numeric
	STAB	stay abroad	no stay, long, short
	YOEAS	years of English at school	numeric
	YOEAU	years of English at university	numeric
contextual variables	TOPIC	topic interviewees chose	country, entertainment, experience & self
	TIME_IN_RECORDING	time in the interaction	early vs. late
	SPEAKING_STYLE	speaking style	mono, dia, both
	INT_SEX	interviewer sex	male vs. female
	INTERRUPTIONS	interruptions by interviewer	numeric
	TOPIC_INITIATIONS	no. of topics initiated	numeric
linguistic variables	IM_LENGTH	length of ips in seconds (med.)	numeric
	IP_LENGTH	length of IPs in seconds (med.)	numeric
	LENGTH	length of ips in seconds (absol.)	numeric
	TUNE_PATTERN	for uptalk tunes	LH, LHH, HH, HLH
	TUNE_PATTERN_confl	tune patterns	HL, HH, LH, LL
	POSITION	position within an IP	accent vs. edge
	PRAGM	pragmatic function per ip	continuation (CON), handing over of turn (HOT), finality (FIN), listing (LIST), emphasis (EMP), checking (CHECK), insecurity (INSEC)
	PRAGM_confl	pragmatic function per ip	continuation, handing over of turn, finality, listing, & other
	QUOTED	quoted speech	quoted vs. non-quoted
	TONE_TYPE	conflated ToBI labels	H*, H-, H%, L*, L-, L%
	UPTALK	uptalk usage	uptalk vs. non-uptalk
fluency variables	UP	unfilled pauses	numeric
	EL	elongations	numeric
	FP	filled pauses	numeric
	HP	hesitation phenomena	numeric
	WPS	words per second	numeric

3.7.3 Statistical tests

The empirical part of the present analysis was carried out with both *Praat* scripts and R-scripts (R Development Core Team 2015) to automatically extract information from the *TextGrids*. Many possible sources of variation were controlled for within the current data analysis (see Section 3.7.2 for an overview). The present study makes use of both descriptive and inferential statistics. First, different visualization methods such as boxplots and scatterplots were used to identify outliers in the data set. All figures have been generated using R-Studio (R Development Core Team 2015). All boxplots and lineplots were made with ggplot2 (Wickham 2016). This method was applied first in order to identify and later correct possible annotation errors (e.g. missing interval mark in a *TextGrid* resulting in a very long IP). In a second step, in order to find appropriate statistical tests, first normality of the data had to be tested. Histograms and residual plots were used to test the normality of the data within each group. According to the central limit theorem (CLT), a sampling distribution of means approaches a normal distribution the larger the sample gets. The CLT thus suggests that a sample of 25-30 is sufficient to produce a normal distribution (Howell 2013: 178-181). Given the large sample size of the present data set, most prosodic features should exhibit a normal distribution around the mean. Despite the large sample size, most prosodic features such as IP length (words and seconds) and f0 are not normally distributed, i.e. skewed to the right. For inferential statistics, logistic regressions, conditional inference trees ctree from the partykit package (Hothorn et al. 2006; Hothorn and Zeileis 2015) and mixed-effects models (lme4, Bates et al. 2015) were employed with speaker status (AmE vs. NBrE vs. CzE vs. GerE vs. SpE (MadE vs. MurE) as a factor and all prosodic measures as the dependent variables. Every dependent variable will be looked at separately with the various independent variables. The modeling of each mixed model and the corresponding factors will be explained briefly at the beginning of the results sections for each prosodic variable. In order to provide an overview of the data, for each prosodic feature separately, I made use of various descriptive statistics such as means, medians, standard deviations, etc. Following Gries' (2013: 45) recommendation on which information to provide in the methods and results section of a thesis, I will always provide the sample sizes, the obtained effect (such as the mean, the percentage, etc.), the name of the test(s) used, its statistical parameters, the p-value, and my decision in favor of or against H1, i.e. the alternative hypothesis. Therefore, this information will be given at the beginning of each results section (except the decision for or against the H1, which will be at the end of each section) together with the statistical models. In sum, the present study employs a multifactorial analysis, of six tones (H*, L*, H-, L-, H%, L%) across two phrase types (ip, IP) and seven regional varieties of English (AmEO CzE, GerE, MadE, MurE, NBrE, and SAmE), amounting to 84 different comparisons. Even more comparisons are made when all other independent predictor variables are added to the respective statistical models (see Section 3.7.2).

3.7.4 Research questions and hypotheses

In the present thesis, I want to test whether learners of English favor certain tones and tunes, and whether they exhibit differences in intonational phrasing, f0 range, and high tunes (i.e. uptalk). My overall hypothesis is that the learners from different language backgrounds will exhibit differences specific for their L1 and thus exhibit features of L1 transfer or -influence in their L2. Based on the statistical research question, certain predictions were made based on findings of previous research, as well as my own personal view based on my extensive analysis of the recorded speech samples. As laid out in Section 1.1.1, the main goal of the thesis is to determine whether there are universal features in non-native speech, which would point towards developmental features of the L2 rather than L1 transfer phenomena. The null hypotheses for all research questions listed in this Section would propose identical values for all varieties of English (all native and non-native ones). The overall alternative hypothesis is that there is a significant difference between the prosodic patterns used by the learners and NSs. The hypotheses include my predictions on the form, function, and distribution of tones, tunes, intonational phrasing, f0 range, and uptalk tunes. All alternative hypotheses (H1) are directional, i.e. (X is greater/smaller than Y), only some of the uptalk-related hypotheses are non-directional, because I did not have any predictions before listening to and analyzing the data. The statistical RQs will be listed in the respective sections in the data analysis (Section 4). The following lists the more specific RQs based on the more general ones formulated in Section 1.1.1, along with the statistic used and the directional hypotheses:

Distributional dimension on tone frequency:

RQ1: *Do learners of English produce more tones than native speakers of English? Does this differ depending on the speaking style, speaker sex, speaker age, fluency phenomena, ip length in seconds, and IP speech rate?*

Statistic used: linear regression

Directional hypotheses: H1: NNSs will produce more tones than NSs

Distributional dimension on tone and tune type frequency:

RQ2: *Do learners of English produce more high edge tones than native speakers? Does this differ depending on REGION, position in the IP, pragmatic function, interviewer influence, topic, SEX, interviewer sex, age, IP length, uptalk, and quoted speech?*

Statistic used: binomial logistic regression

Directional hypotheses: H2: NNSs will produce more high edge tones than NSs and will therefore have more tunes in a higher register.

Distributional dimension on IU frequency:

RQ3: *Do learners of English produce more prosodic breaks (IPs/ips) than native speakers of English? Does this differ depending on the speaking style, and fluency (measures in WPS, number of silent and filled pauses, number of hesitations)?*

Statistic used: linear regression

Directional hypotheses: H3: NNSs will produce more IPs/ips than NSs

Realizational dimension on IU length:

RQ4: *Do learners of English produce slower speech rates in IPs/ips compared to native speakers? Does this differ depending on speaker sex, age, speaking style, and fluency?*

Statistic used: linear mixed effects model

Directional hypotheses: H4: NNSs will produce a slower speech rate in IUs

Realizational dimension on f0 level:

RQ5: *Do learners of English produce native-like f0 levels? Does this differ depending on the speaking style, tune pattern, sex, age, number of topic initiations, position in the IP, ip length, pragmatic function, number of interruptions, and hesitations?*

Statistic used: linear mixed effects model

Directional hypotheses:

H5: F0 level measures of NNSs will be closer to the TL norm
H6: F0 level differences will be position-sensitive
H7: F0 level will differ by ip length
H8: F0 level will be higher in monologues
H9: F0 level will be higher in a checking than in handing over turn
H10: F0 level will be higher in post-interactions
H11: F0 level will be higher the more topics are initiated
H12: F0 level will be higher the more hesitations are produced, because restarts and many smaller ips might contribute to pitch resetting (cf. Hincks 2004)

Realizational dimension on f0 span:

RQ6: *Do learners of English produce a narrower f0 span than native speakers? Does this differ depending on the speaking style, tune pattern, sex, number of topic initiations, position in the IP, ip length, and pragmatic function?*

Statistic used: linear mixed effects model

Directional hypotheses:

H13: F0 span measures of NNSs will be narrower than those of NSs
H14: F0 span differences will be position-sensitive
H15: F0 span will be wider in longer ips
H16: F0 span will be wider in dialogues
H17: F0 span will be wider in checking than in handing over turn
H18: F0 span will be wider in post-interactions
H19: F0 span will be wider the more topics are initiated
H20: F0 span will be wider the more hesitations are produced

F0 range and entrainment:

RQ7: *Do native and non-native speakers entrain their f0 range to that of their respective interviewers?*

Statistic used: linear mixed effects model

Directional hypotheses:

H21: NSs and NNSs will both entrain to their respective interlocutors

H22: NNSs will entrain more to their interlocutors than NSs
H23: Female speakers will entrain more to their interlocutors than male speakers

Distributional dimension on uptalk frequency:

RQ8: *Does the distribution of uptalk differ depending on the region, speaking style, age, topic, sex, interviewer sex, ip length, and the pragmatic function used?*

Statistic used: binomial logistic mixed effects model

Directional hypotheses:

H24: Female speakers will use uptalk more frequently
H25: Uptalk tunes will be produced more frequently in monologues
H26: The longer the utterances, the more uptalk instances will be used

Realizational dimension on uptalk tunes:

RQ9: *Does the f0 level and span of uptalk differ depending on the region, speaking style, age, topic, sex, interviewer sex, position in the IP, and pragmatic function used?*

Statistic used: linear mixed effects model

RQ10: *Are there differences in the relative slope of the rise in uptalk tunes in the different varieties of English? Does this differ depending on the tune shape, speaking style, age, interviewer sex, position in the IP, ip frequency, and pragmatic function?*

Statistic used: linear mixed effects model

At this point, there is little or no previous research on uptalk that would permit any conclusions on the shape and function of uptalk in non-native speech. While annotating the data, smaller differences were perceived by the author of the present thesis between the uptalk tunes of the different learner groups. This impression, however, needs to be statistically tested. Therefore, no directional hypotheses are formulated for RQ9 and RQ10.

The stay abroad variable:

While the stay abroad variable and RQ11 were formulated before the data analysis was conducted, i.e. RQ11: *Does the overall prosodic performance of the learners improve after a stay abroad in an English-speaking country? Does this differ depending on the length of stay abroad (STAB), country of stay abroad (COUNTRY), learner variety (REGION), and speaker sex (SEX)?* More specific sub-questions were added after the analysis of each prosodic feature was conducted:

RQ11a: *Do learners produce more or fewer tones after a stay abroad?*

Statistic used: linear regression

RQ11b: *Do learners produce fewer high edge tones after a stay abroad?*

Statistic used: binomial logistic regression

RQ11c: *Do IU breaks become less frequent after a stay abroad?*

Statistic used: linear regression

RQ11d: *Do learners produce a faster speech rate (WPS) per IU after a stay abroad?*

Statistic used: linear mixed effects model

RQ11e: *Do learners produce a lower f0 level for LH- and HH-tunes after a stay abroad?*

Statistic used: linear mixed effects model

RQ11f: *Do learners produce a wider f0 span for HL-tunes and narrower span for LH- and HH-tunes after a stay abroad?*

Statistic used: linear mixed effects model

RQ11g: *Does uptalk become less or more frequent depending on the length of stay abroad and the variety of English the learners have been exposed to?*

Statistic used: binomial logistic mixed effects model

RQ11h: *Do uptalk f0 measures change after a stay abroad and the variety of English the learners have been exposed to?*

Statistic used: linear mixed effects model

These hypotheses are tested in Section 4 and revisited in the discussion in Section 5.

3.8 Caveats, solutions, and discussion

The caveats of the present study mainly lie in some limitations of the corpus data (corpora that were not primarily compiled for purpose of phonetic analyses), as well as general issues with annotating and analyzing prosodic variation. While the LINDSEI and LOCNEC corpora are an invaluable resource for spoken discourse studies, they, however, were not created with prosodic studies in mind. Shortcomings of the LINDSEI corpora are, as mentioned elsewhere (Gráf 2015), for instance that there are more female than male participants, differences in task execution (due to fuzzy instructions), more metadata could have been collected (e.g. proficiency levels, IQs, musicality, more detailed learning context variables, purpose for stay abroad, regional origin of speakers, personality test, motivation, attitudes, prosodic training experience, etc.), and higher quality recordings could have been made. Additionally, for transfer studies it is indispensable to have recordings of the same learners available in their native languages. Such recordings were not available[72] for the learner corpora. L1 influence in the present study can only be determined on the results of previous studies on the respective L1 prosody. Also developmental SLA aspects cannot be answered satisfactorily, since

[72] However, LINDSEI-CZ, which was a corpus with excellent sound quality, now also includes recordings of the same learners fulfilling the same tasks in Czech. These files are, however, not publicly available.

longitudinal data is not available. However, a pseudo-longitudinal approach (see Götz and Mukherjee 2018) was undertaken with the present data set. Despite these limitations of the selected corpora, they still include many advantages to the present study in terms of the objectives it tries to achieve (see Section 1.1.1). The corpora are a good choice for the present study, because they contain objective language data, which comes closer to authentic and natural language use in spontaneous contexts students of English would encounter in real-life. Also, some of the NS corpora (LOCNEC, NSV) used in this study have not been used for prosodic analyses to the authors' knowledge (except the NWSP corpus). Therefore, these corpora are virtually unexplored in terms of prosody and this could be nicely paired with all the previous research in other linguistic areas.

A question to be asked is whether the learner samples analyzed in the present study are applicable or generalizable to the entire population of (advanced) learners, as well as whether they are comparable to the NSV speakers who present a group of non-students. It can be argued that university students may not be ideal candidates to represent entire learner populations, because they range above average for most variables such as intelligence and motivation. However, previous research (Hollien et al. 1997: 2989) found that 157 male university students compared to a group of males in the US military (n=142) found the mean f0 of university students to be higher by about 8 Hz (129.4 Hz vs. 121.5 Hz) and over 1 ST difference between the two groups, which was a statistically significant difference but "well within the expected f0 range for young adults" (2990). Therefore, if there are any differences between university learners and learners outside of the university, these should be rather small.

Additionally, the "advanced" learners in my study are far from being a homogenous group. Especially the SpE speakers, while having an L1 in common, they represent two distinguishable groups. The different proficiency levels in the LINDSEI corpora, contrary to the advancedness claims made by Gilquin et al. (2010), the corpus represents a useful resource that comes close to the actual population a teacher may find in his/her classroom, who have different needs and problems in language learning. However, caution has to be taken when generalizing to results of the present study to lower (B1-B2) and higher (C2) proficiency levels, because most of the learners are or can be judged as representing a C1 proficiency level. Another drawback is that the current corpus does not include proficiency levels of one of the learner groups (LINDSEI-SP and half of LINDSEI-GE) and some learners in each of the sub-corpora seem to be more advanced or fluent than others. According to Gilquin et al. (2010), the students recorded in the LINDSEI corpora could be placed in the B1-C1 range (intermediate to advanced level) on the Common European Framework scale. Huang et al. (2018) and Huang (2019), who conducted professional posthoc CEFR ratings on all LINDSEI-CZ files and 20 LINDSEI-GE files, indeed show that most of the learners are placed on the C1 level, and very few on the B2 and C2 levels. According to Pérez-Paredes and Sánchez Tornel (2015: 158) "the LINDSEI-ES informants cannot be considered as advanced learners of English". One possible solution to compensate for the missing CEFR ratings is to take the institutional status and amount of time in an English-speaking country as possible indicators of proficiency for those learners whose CEFR levels are not available. In addition, the present study will test the available CEFR levels with the fluency MLR results from Götz (2013) and Gráf (2015) and with lexical complexity measures. This "seeming weakness", as Gráf (2015: 123) puts it, can also be seen as an opportunity which allows us to look at different proficiency levels of learners in order to determine whether even the lower proficiency groups share certain prosodic traits with the more advanced ones. Analyzing a more heterogeneous group in terms of proficiency is also closer to the actual reality of the population one would find in a language learning classroom or university setting.

The learners' proficiency levels are not the only caveat; the level of formality of the recording situations also needs to be discussed. Mukherjee (2001) found that the lower the level of formality and the higher the degree of planning, the higher the length of IPs. Tyler (2019) found that informal conversations have larger rise spans. Interviews, which are quite often used as a source for intonational studies, suffer from the problem of the "Observer's Paradox" (Labov 1966: 99; 1972: 90), which has to be accounted for. The presence of a recording device and an interviewer, who might in some cases be the lecturer of the interviewee, all contribute to a perceived formality, which, in turn, might hinder casual and naturally occurring conversation (cf. Rüdiger 2016: 66). Furthermore, there are types of frames that the interviewee chooses over which the researcher has no control (cf. Schilling 2013: 128). For instance, there are interviewees who speak more freely and even ask the interviewer questions, even though this was not required in the instructions, compared to others who strictly follow the interview situation rules and the interviewer's instructions. These interview rules involve who might say what, the question-answer format, only the interviewer introduces a new topic, etc. (Wolfson 1976: 192). Additionally, when norms of speaking are uncertain or violated, then 'unnatural' speech might occur (ibid.: 202).

In the present study, these two functions will also have to be investigated by separating more emotionally loaded speech from more factual or neutral referential speech, both being further classified into negative and positive experiences. To give examples, one of the Czech learners (CZ004) speaks about a very pleasant and emotional experience she had with her father on a trip to London, while speaker CZ002 speaks about his passion about theater plays. Although there are some speakers who had a very bad experience and report it in an emotional manner, others might report emotional events in a rather referential and factual manner (e.g. CZ017). However, most speakers in the corpus analyzed in the present study rather exhibit referential functions in their speech samples rather than emotional speech. All of the 225 files in the current corpus were recorded in a university setting and they can be described as spontaneous speech with varying degrees of (perceived) formality, depending on interlocutors involved in the recordings. While some of the interlocutors are a bit more familiar with each other than others, they are all in a student-lecturer relationship, except the equal status conversations of the NSV corpus. The individual corpus files selected for the present study can include either purely monologic speech, dialogic informal interviews, or a mix of both. Another factor that can give an insight on the degree of formality is the distribution of quoted speech utterances. Quoted speech occurs more often in more informal and intimate interactions and also includes the largest rise spans (cf. Tyler 2019). Therefore, the quoted speech samples will be explored to see whether a more formal or informal speaking style characterizes the present data set.

Another issue that has to be addressed is the selection of speech samples for each of the 225 audio files. Resulting from this selection procedure of the most fluent part of a speaker's larger sample of speech, possible prosodic patterns might have been obscured and the results of the analysis might be attributed to the way the data was annotated. Thus, while one speaker may make use of many uptalk instances in the 560 words annotated, another speaker might not exhibit any uptalk instances, or just a few. Therefore, when uptalk instances are found or non-existent, this might depend on the part that was selected for annotation. Despite this selection problem, some speakers may make sporadic use of uptalk in general, while individual speakers may be responsible for the high frequency of uptalk instances in the respective group. This further complicated the modeling of the data. Future research on this data set could include the annotation of the entire recordings or the annotation of samples from the same speaker(s) at the beginning and at the end and oppose them to see whether speakers would be characterized as different speakers or whether the prosodic features persist. Due to these differences in audio quality and fluency of individual speakers, the distribution of speaking styles (mono and dia) is not distributed evenly among the corpora. The most balanced corpus in terms of speaking style is the NSV corpus.

Another issue in the present study is that intonation is highly context dependent and the corpus created for the purpose of this study has seemingly infinite contexts. Nevertheless, this could at the same time be a highly diverse and insightful resource for L2 prosody research. As Rüdiger (2016: 53) states, "all interactions are prone to bias and context-dependency". Speaking styles such as interviews and other speaking styles, for instance, contain various speech events, which again, fulfill different functions, such as responding to interview questions or telling a story, and thus there are different types of text that are being produced even within one and the same genre (cf. Warren 2016: 124). The recordings used in this study were collected under slightly different conditions. Ideally, both the LOCNEC and LINDSEI interviews were conducted by NSs of English interviewing native and NNSs of English. However, in some cases the interviewers are not NSs of English but of other languages. This applies to all LOCNEC files and half of the LINDSEI-CZ files. Furthermore, the interviewees would ideally receive the choice of topics shortly before the interview and they were not allowed to take any notes, as it was supposed to represent spontaneous speech (cf. Brand and Kämmerer 2006: 132). This, however, is not the case for most sub-corpora of LINDSEI. For example, the Spanish students received the choice of topics after they were asked for their personal information and while the recording was running. The German learners, for example, had a little bit more time to prepare before the recording started. The Czech learners were given a few minutes to prepare for the interview (Gráf 2019: 179). The 225 files of the present study all consist of spontaneous speech samples. By spontaneous speech I refer to Barry and Andreeva's (2001: 55) definition of "unscripted, unprepared in terms of the number, organization and expression of the information points it communicates". However, the data set cannot only be classified as "spontaneous speech" but also as interactional talk. So, within the overarching speaking style of spontaneous speech, one can further distinguish between monologic and dialogic speech. Choosing spontaneous speech data over read speech samples opens up the possibility of analyzing more authentic prosodic patterns. However, it is not as directly comparable as read speech, where utterances can be directly compared to each other by different speakers and where intonational phrasing is controlled for. In spontaneous speech, intonational phrasing is unpredictable and so is the number of tonal patterns. Nevertheless, one major advantage of including spontaneous speech over read speech is that in language assessment usually informal interviews or shorter monologues are employed to assess the learner's spoken ability. However, prosody is usually not assessed by any teacher or practitioner. Therefore, the present study could help to characterize spoken interactions between interlocutors and provide further empirical evidence of prosodic deviance that should be attended to in language teaching and testing.

Another problematic aspect that has to be mentioned is that of dialectal variation, specifically the GA dialect, which is said to be spoken generally by speakers from the Midwest and Midlands of the U.S.A. The current study mainly includes speech samples from Southern speakers of the U.S. It is often stereotyped and sometimes also confirmed that SAmE is articulated slower and exhibits different prosodic patterns. The learners of English in the present study, however, have been spread out all over the U.S., e.g. in Texas, Florida, New York, Wisconsin, Minnesota, etc. While it can be argued that SAmE might not be the most representative dialect of "standard" AmE and might not be spoken by the learners in the corpus, it can, nevertheless, be said that even Midland dialect speakers share many prosodic features with Southern speakers (cf. Clopper and Smijlanic 2015) and, in more formal tasks, regional features might diminish or disappear completely. For that reason the NWSP recordings were added, which include further AmE dialects.

Additionally, which pragmatic functions and emotions a speaker experienced or intended cannot always be measured with absolute certainty, and has to be inferred from the discourse (cf. Patterson 2000: 50). Patterson (2000: 18) contends that a combination of prosodic features such as pitch, speech rate, rhythm, and loudness contribute to the expression of emotion in speech. Even though most of the speakers in the present study do not represent an emotional speech trend, the few that do have instances of emotional speech were annotated as such. This will make an analysis of individual variation even more detailed, even though a detailed annotation of the acoustic correlates of emotion was not undertaken in the present study. Merely the fact that emotional speech might be present in the annotated sample was noted. While there are some speakers who had a very bad experience and report it in an emotional manner (e.g. one of the German learners who had a very bad experience with an American girl at a party who invited men to their house), then there are others who also had bad experiences (e.g. CZ017 who fires up a fire cracker and injures himself very badly and describes his stay at the hospital and his parents reaction), which are reported in a very referential and factual manner despite the emotionality of the topic.

Finally, besides the issues of working with spontaneous speech from corpora that have not been primarily collected for phonetic analyses, I would like to point out some general problems with the annotation of prosodic material. As Jilka (2007: 78) points out, the chosen model of intonational description shapes the way that intonational features are perceived and ultimately annotated. Additionally, "[…] it cannot be determined with absolute certainty which model if any reflects the true representation of an intonational phenomenon" (Jilka 2007: 78). The decision which label should be given to tunes in the f0 patterns is difficult to agree on and is limited by the descriptions of the chosen annotation scheme. Prosodic annotation is extremely impressionistic in nature and subjective. Therefore, the present study made use of different procedures to ensure more objectivity in the annotation of the data set (see Sections 3.5 and 3.6). The ToBI (Silverman et al 1992; see Section 3.4) annotation system has been chosen as the annotation standard for the present study, because it is most widely used in current and past intonation research. The limitations of ToBI and how they were dealt with in the present study are discussed in Sections 3.4 and 3.7.1.1. Coding consistency is another critical issue which is hard to maintain. In order to compensate for this issue, a detailed manual (see Appendix 1) has been set up and weekly meetings were conducted to discuss difficulties and dubious cases together.

3.9 Summary

The present study adopts a mixed methods approach (corpus-based, corpus-driven, instrumental, auditory, and quantitative) and a multivariate analysis in examining native and interlanguage data. This study sets out to characterize the prosodic features produced by three L2 learner groups (CzE, GerE, and SpE), and investigates the extent to which the learners adopt native values of the TL (BrE and AmE). Within the autosegmental-framework this study reports on a study on L2 learners' intonational deviances in spontaneous monologic and dialogic speech derived from a CIA and CA (cf. Granger 1996) of Czech, German, and Spanish. Through quantitative analyses, the structure and function of different prosodic features are compared. The prosodic features analyzed are the following: tones and tunes, intonational phrasing, f0 range, and uptalk. The interlanguages based on the Czech, German, and Spanish components of the LINDSEI (Gilquin et al.2010) are compared to English native speech with prosodically annotated versions of the LOCNEC (De Cock 2004), representing BrE, and the NSV (Atkins 2017) and the NWSP (Clopper and Pisoni 2006) corpora, representing AmE. Since neither of the corpora include prosodic annotations, the manual annotation had been performed with the ToBI (Silverman et al. 1992) system and *Praat* (Boersma and Weenink 2019). The corpus sample in the present study consists of speech from a total of 225 speakers, with 135 learners varying in age, gender, L1, proficiency, length of stay abroad in English-speaking countries, and different interview situations. In order to ensure data comparability, the corpus contains similar spontaneous speech on similar topics (a country travelled to, an (childhood) experience made, a movie or play) produced in an interview situation, the same age group (18-33, 22 years on average), and same length of speech (≈560 tokens=1.9-5.9 minutes, on average=3.13 minutes) consisting only of declaratives from learner and native speech. For the analysis of the files, information was extracted and significance tests were conducted with the help of *Praat*-scripts (Boersma and Weenink 2019) and R-scripts (R Development Core Team 2015). In this context, the present study provides new insights from a descriptive perspective, i.e. by including an analysis of prosodically virtually unexplored corpora from three different learner varieties from three different language families and the neglected speaking style of spontaneous speech. Secondly, the present study also adopts a different methodological perspective by including various extralinguistic variables and complex statistical models.

4. Data analysis of prosodic features in the different varieties of English

This section reports on the results of the phonetic and phonological multifactorial analysis of the four prosodic features (tones and tunes, intonational phrasing, f0 range, and uptalk). For each feature, descriptive and inferential statistics are provided. Since the main goal of the present study was to investigate what type of prosody L2 learners of English produce in declaratives and how their productions compare to those of NE speakers, the following steps will be taken:

1) the productive prosody features are considered in isolation
2) these features are then compared to the native control's performance in order to determine deviances and target-norm approximations of individual performances
3) intra- and intergroup variations are investigated

The results section is followed by a discussion of all features regarding the intonational dimensions and research questions.

4.1 Tones and tunes in native and non-native speech

The present section starts with a birds-eye view on the tones and tunes that occur in the corpus data. Descriptive statistics will be given first and will then be complemented by inferential statistics. The frequency of tones and tunes will be described by all predictor variables, e.g. REGION, SEX, AGE, SPEAKING_STYLE, different fluency variables (HP, FP, UP, and EL), and PRAGM to account for any pragmatic effects. In order to determine the overall tone frequency, the rate of use of H and L tones and tunes across the different L1 populations, various models were built. The statistical tests address the research question pertaining to structural and functional differences in tones and tunes in native and non-native speech. As aforementioned, the different varieties of English will be referred to as: Northern British English (NBrE), Southern American English (SAmE), American English Other (AmEO), Czech English (CzE), German English (GerE), Madrid English (MadE), and Murcian English (MurE) and summarized under the IV REGION. The predictor variables involve different sociolinguistic variables (e.g. sex and age), contextual variables (e.g. interview variables), and linguistic variables (e.g. prosodic breaks/intonational phrasing, fluency features).

4.1.1 Tone frequency

First the tone frequency will be described in general to answer RQ1 concerning whether learners use more tones and which factors determine tone frequency. Then the distribution of tones and their ToBI labels will be considered by position (Sections 4.1.2-4.1.5) and how they combine into tunes (4.1.6). In order to answer RQ2, which asks whether learners use high edge tones more frequently, especially in IP-final positions. Again, other factors such as SPEAKING_STYLE, AGE, SEX, REGION, and fluency phenomena (elongations, filled and silent pauses, and hesitations) will be tested to investigate how the distributions can be explained. While dysfluencies are the norm even in native speech, the occurrence of hesitations and filled pauses may lead to the production of extra accents in native speech, while NSs may pause and hesitate without producing extra accents (cf. Rasier and Hiligsmann

2007: 58, with L1/L2 Dutch and French speakers). The IP length (in seconds) will be included as a factor as well, since previous research has determined that differences between native and non-native speech become more pronounced in longer stretches of speech (Ramírez Verdugo 2006b: 524). First, however, the overall descriptive statistics of the distribution of tones across the regional varieties of English are given in Table 15. The REGION variable includes seven levels, i.e. AmEO (including AmE from the North, West, Midlands, and Mid-Atlantic), CzE, GerE, MadE, MurE, NBrE, and SAmE. These descriptive statistics were calculated with "stat.desc()" from the package "pastecs" (Grosjean and Ibanez 2018).

Table 15. Descriptive statistics for tone frequency

Measures of central tendency	**NBrE**	**SAmE**	**AmEO**	**CzE**	**GerE**	**MadE**	**MurE**	Measures of disp. in brackets
total number of tones (absolute and normalized phw)	8,978 1,603	6,358 1,133	3,028 542	8,892 1,587	9,894 1,765	5,677 1,005	4,616 920	
no. of tones (phw) by sex	f: 871 m: 732	f: 783 m: 350	f: 209 m: 333	f: 1,377 m: 210	f: 1,283 m: 482	f: 804 m: 201	f: 668 m: 252	
median freq. of tones by region (absolute vs. normalized freq. in phw)	201(27) 36 (4)	228 (42) 41 (7)	190 (24) 34.5 (5)	200 (36) 35 (6)	214 (39) 38 (7)	229 (34) 40 (7)	233 (48) 47 (7)	**IQR**
min-max number of tones (absol. and norm. freq.)	150-233 26-42	163-270 29-49	149-229 27-40	146-246 26-44	140-304 25-54	181-270 33-48	119-307 36-59	
median no. of tones per IP/ip	7 (2) 2 (1)	8 (3) 2 (0)	7 (2) 2 (1)	8 (1) 3 (0)	7 (1) 3 (0)	8 (1) 2 (0)	7.5 (1) 3 (1)	**IQR**
min-max no. of tones per IP/ip	5-11 2-3	4-15 2-3	5-14 2-3	6-13 2-3	5-10 2-3	6-10 2-3	5-11 2-3	

If normalized and logged values are considered, only the MurE speakers seem to have a tendency to produce tones more frequently than the NS groups. The boxplot in Figure 54 (see Appendix) visualizes the results. Tone frequency depends heavily on intonational phrasing and vice versa. The smaller the intonational units are, i.e. the more ips there are, the more tones there are in all corpora. This is because ToBI annotations require that there are at least two tones (*, -) per ip, and at least three IP-finally (*, -, and %). From Table 15 one can see that the number of tones per IU is quite similar across all corpora. Also the longer the phrases are in seconds and words (stronger correlation), tones may become more frequent. These descriptive statistics are complemented by inferential statistics in order to answer RQ1: *Do learners of English produce more tones than NSs? Does this differ depending on the speaking style, speaker sex, speaker age, fluency phenomena, ip length in seconds, and IP speech rate?*

H0	the distribution of the DV TONES_PHW of the NNSs equals that of the NSs, and does not differ depending on the levels of the predictors (IVs & their interactions): REGION, SEX, AGE, SPEAKING_STYLE, FP, UP, HP, EL, IM_LENGTH, and WPS_IP: $R^2 = 0$.
H1	the distribution of the DV TONES_PHW of the NNSs is larger than that of the NSs, and differs depending on the levels of the predictors (IVs & their interactions): REGION, SEX, AGE, SPEAKING_STYLE, FP, UP, HP, EL, IM_LENGTH, and WPS_IP: $R^2 > 0$.

A linear regression was selected for this research question because there was only one data point per speaker, i.e. the specific number of tones per speaker. Therefore, in order to address RQ1, a linear regression with backward model selection was employed with REGION, SEX, AGE, SPEAKNG_STYLE, fluency features (filled pauses, elongations, hesitations, and WPS per IP), and ip length as IVs and TONES_PHW as DV. The predictor variables were tested for multicollinearity and the data was also tested for overfitting; both tests were negative. NBrE was introduced as the intercept, because its corpus design is the most comparable to the learner files and it will be used as an intercept for all statistical models that include comparisons with a NS norm. The following final model ($R^2=0.78$, $F = 68.97$ on 12 and 212 DF, $p<0.001$***) was fit and the results are represented in Table 16. No interactions of the various factors significantly improved the model fit. The results are additionally visualized in the following effects plots in Figure 11:

Table 16. Results for linear regression with tones phw. Final model: TONES_PHW ~ 1 + REGION + SEX + SPEAKING_ STYLE + WPS_IP + UP_NORM + IM_LENGTH

Predictors	Estimate	Std. Error	t-value	p-value	
(Intercept)	63.2384	5.6404	11.212	<0.001	***
REGION_SAmE	1.0081	0.6907	1.46	0.1459	
REGION_AmEO	-0.3492	0.793	-0.44	0.6601	
REGION_CzE	0.9939	0.8594	1.156	0.2488	
REGION_GerE	1.1758	0.7036	1.671	0.0962	.
REGION_MadE	-1.1856	0.8164	-1.452	0.1479	
REGION_MurE	1.0104	0.9697	1.042	0.2986	
SEX_male	-1.8139	0.4037	-4.493	<0.001	***
SPEAKING_STYLE_both	1.1747	0.5423	2.166	0.0314	*
SPEAKING_STYLE_dia	1.1254	0.5025	2.239	0.0262	*
log2(WPS_IP)	-12.9266	1.446	-8.939	<0.001	***
log2(UP_NORM)	3.1964	0.7145	4.474	<0.001	***
IM_LENGTH	-9.3158	0.7154	-13.022	<0.001	***

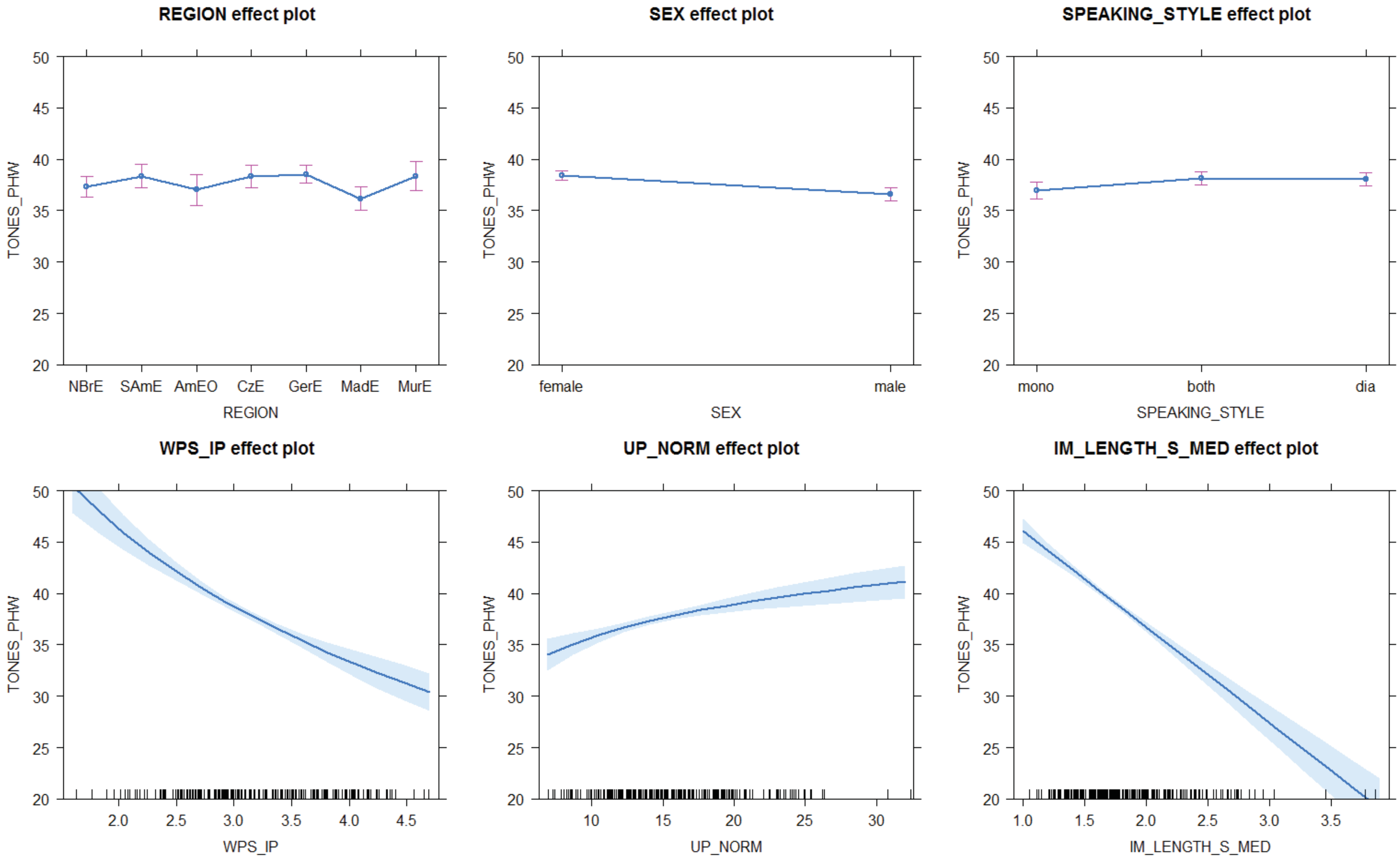


Figure 11. Effect plots for linear regression results of tones phw

NBrE females in a monologic speaking style are predicted to use 63 tones phw. The model results show that when it comes to the IV REGION, there are no significant differences between any of the varieties of English (native and non-native) (see Plot 1 in Figure 11)[73]. Although all learners (and SAmE speakers) exhibit the tendency to produce one tone phw more than the NBrE intercept, except MadE speakers who tend to produce one tone phw less. The model results also show that female speakers are predicted to use more tones phw than males (-1.81, $p<0.001$***) (Plot 2). Additionally, the more dialogic speaking styles (both and dia), the more tones are produced, or the speech becomes more prosodically diverse (i.e. more lively and possibly entrained by the interviewer) compared to the monologic parts (Plot 3). Furthermore, the faster the overall IPs in WPS the fewer tones are produced (-12.93, $p<0.001$***) (Plot 4). The more filled pauses are produced, the higher the number of tones phw (3.20, $p<0.001$***), because each pause potentially leads to another ip, which by default has to include at least two tone labels if it is utterance-medial (i.e. a pitch accent and a phrase accent) and three tones if it is utterance-final (i.e. pitch accent, phrase accent, and boundary tone) (Plot 5). The longer the ips (median ip length in seconds) the fewer tones are produced (-9.32, $p<0.001$***) (Plot 6). The predictor variables, SEX ($p<0.001$***), SPEAKING_STYLE ($p=0.04852$*), IM_LENGTH ($p<0.001$***), UP_NORM ($p<0.001$***), and WPS_IP ($p<0.001$***) are factors that explain the variation better than REGION ($p=0.01018$*). Thus, female speakers are predicted to produce more tones, dialogic speech contains more tones, the more silent pauses are produced, the more tones are produced, and longer ips and faster speech rate per IP lead to a lower production of tones phw. In sum, for the variable REGION, H0 has to be accepted for RQ1.

A second model ($R^2=0.75$, $F=24.98$ on 14 and 100 DF, $p<0.001$***) contained only the complete data of the CzE, GerE, and MadE speakers and all learner variables, i.e. the STAB variable (no stay abroad, short stay abroad, and long stay abroad), years of English at school and university, the number of other foreign languages they speak, the age of the learners, as well as the variables from the analysis above to determine predictions on tone frequency. If the learners are compared to each other (complete data only) with all the learner variables, determined that the same IVs (WPS_IP, UP_NORM, IM_LENGTH, and SEX (see Plots 2-5 in Figure 12) play a significant role in predicting how many tones will be used. The only learner variable that significantly improves model fit was the AGE variable, with older learners producing more tones (0.34, $p=0.009176$**) (Plot 1 in Figure 12). None of the other learner variables significantly improved model fit. Also, the interaction between REGION:STAB did not yield any significant differences between the learner groups. Therefore, for RQ11a (*Do learners produce more or fewer tones after a stay abroad?*), it can now be said that there is no significant change after a stay abroad for tone frequency measured in tones phw, all other predictor variables being accounted for.

[73] The plots here and in the following sections are numbered from left to right line-wise.

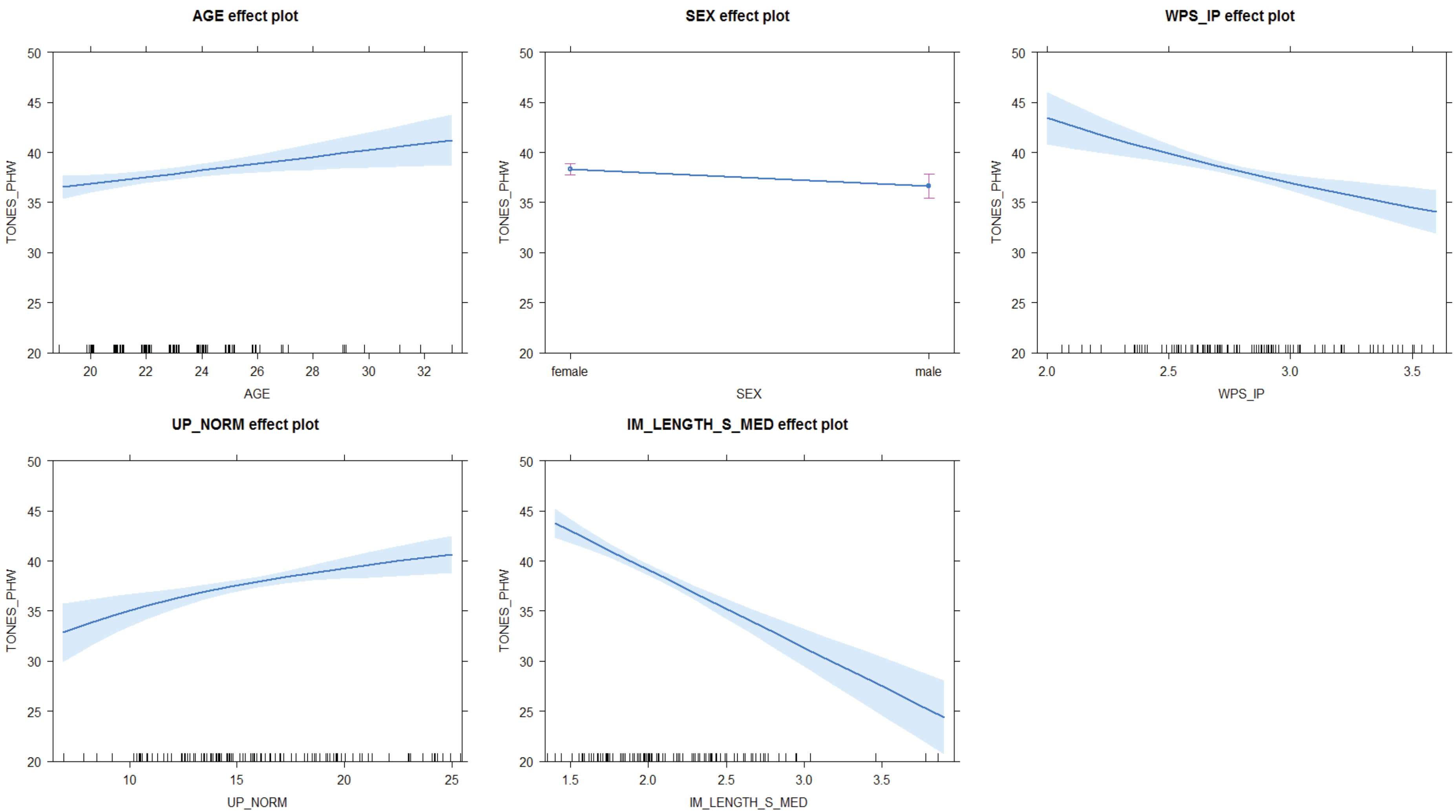


Figure 12. Effect plots for linear regression results of tones phw for the complete learner data only (main effects only)

4.1.2 ToBI label analysis

After the tone frequency has been determined, the present section sets out to provide insight into the second hypothesis that NNSs produce more high edge tones than NSs. This section shows the results of the ToBI labeling of IUs. First descriptive statistics and then inferential statistics will be provided. The tone label frequencies are visualized in Figures 55 and 56 (see Appendix). The initial high boundary tone (%H) is equally infrequent in all varieties of English. The relative frequencies show clearly that all learner groups deviate the most in producing more high phrase accents (H-) than NSs, and consequently L- productions are lower. While all learners produce fewer low boundary tones (L%) compared to the NSs, the CzE and GerE speakers produce the most high boundary tones (H%). With the pitch accents different groupings arise. While the MadE speakers resemble the NBrE speakers in both high (H*) and low (L*) pitch accent distributions, the MurE and CzE speakers produce the most high (H*) pitch accents, and the GerE speakers are more like the AmE speakers with regards to high (H*) and low (L*) pitch accents. Table 17 shows that low tones are a lot more frequent overall as edge tones (L-, L%,) and high tones are a lot more frequent as pitch accents (H*) in ip-medial position. This is the general trend across all varieties of English, except the CzE, where it is almost 50/50 between a high and a low boundary tone.

Table 17. Descriptive statistics, absolute (sum) and relative frequencies (mean) of high and low tones by REGION and POSITION (accent vs. edge) of the reduced data set

	pitch accent				**edge tones**			
	low		**high**		**low**		**high**	
NBrE	838	9.29	3,099	34.63	3,935	21.86	1,107	4.11
SAmE	1005	15.60	1,873	29.60	2,726	21.43	754	3.98
AmEO	387	12.84	971	32.04	1,320	21.99	350	3.71
CzE	1,159	13.27	3,389	37.98	2,006	11.17	2,338	8.80
GerE	1,756	18.01	2,785	28.06	3,066	15.57	2,287	7.59
MadE	589	10.27	1,966	34.79	1,980	17.37	1,142	6.73
MurE	348	7.34	1,777	38.63	1,525	16.46	966	7.03
Total	6,080		15,860		16,558		8,944	
Total instances analyzed:							47,442	

These overall findings are further visualized in the ctree in Figure 13, which shows that a simple distinction between regional varieties of English is highly significant with NSs on the right side and all NNSs on the left. What the NSs all have in common is that a high-low (H* L-L%) pattern is the most common pattern. However, there are significant differences in that SAmE speakers produce fewer high pitch accents (H*) and more low pitch accents (L*) instead. The AmEO group follows a similar trend, while the NBrE speakers strictly stick to the high-low (H* L-L%) pattern. The learners, on the other hand, use the tones on the higher spectrum more often. While the H% is infrequently used in native and non-native speech, it is used more by the CzE and GerE speakers. Compared to the natives, the learners use high edge tones (H- and H%) more often, and as a consequence low edge tones (L- and L%) are used less. Thus, learners also follow a H* L-L% pattern most of the time, but they also produce significantly more instances of H* H-H% or H* H-L% tunes.

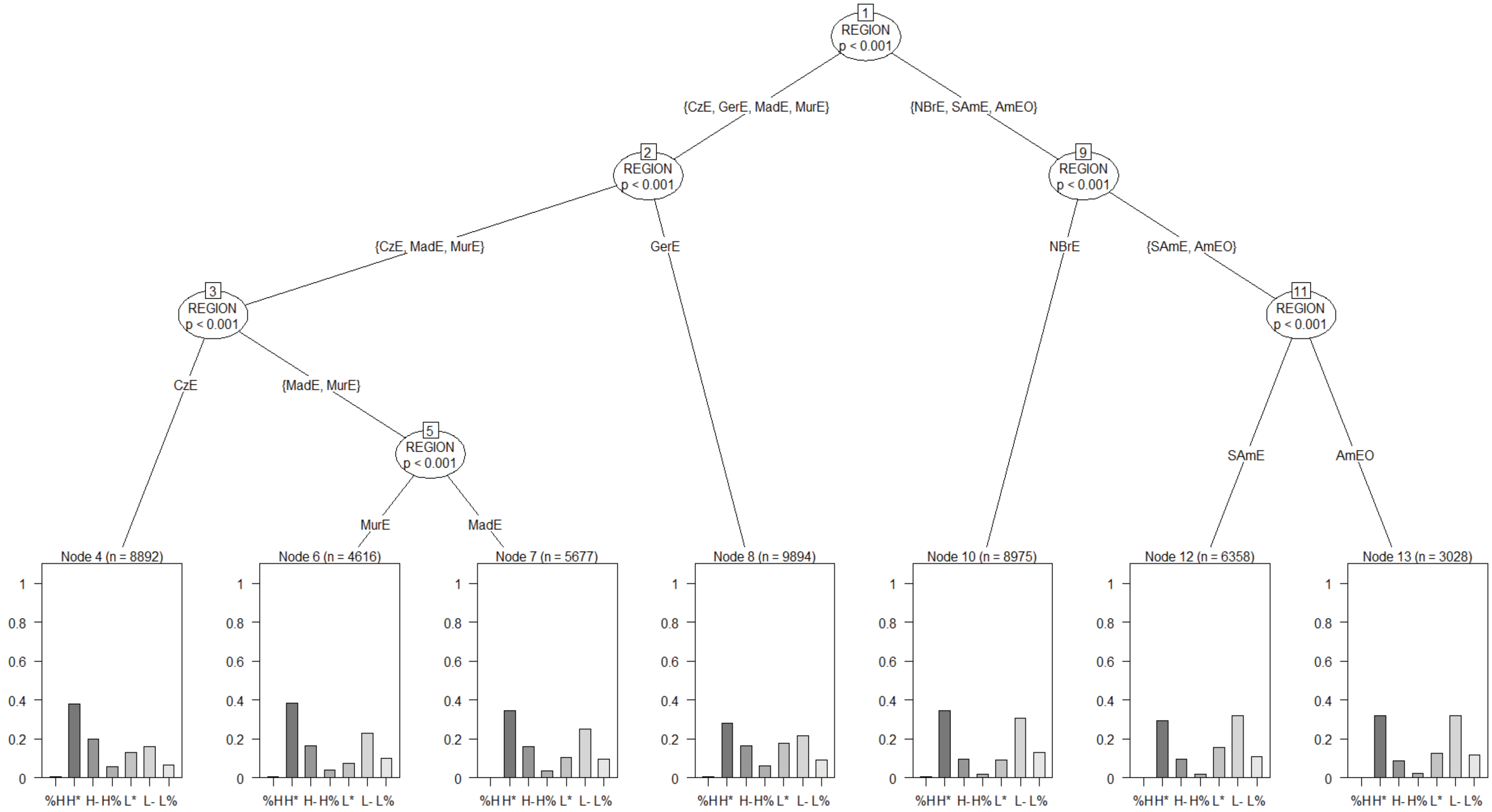


Figure 13. Ctree for ToBI label by REGION (based on absolute frequencies and conflated tones)

The focus of the following paragraphs will be to discuss whether the distributions of tones are significant and which factors determine the choice between high and low tones by their position. Since the outcome variable (H or L) is binary, a logistic regression was performed to test which of the two levels is predicted to be used more if other contextual variables are accounted for. In order to conduct a logistic regression, a few modifications to the data set had to be undertaken. First, the high initial boundary tone %H was removed from the inferential analysis, because in comparison to the frequency of the other tones, there are only very few instances of that tone (n=237), and it is the only tone where only a high tone is possible, i.e. no binary distinction is possible for this category. All other positions: star (*), minus (-), and percent (%) as a final boundary tone) can consist either of a low or high tone. In addition to that, the edge tones (minus (-) and percent (%)) were conflated to one category for the inferential analysis, because a distinction between the two positions (-, %) lead to vif-values of over 5, which are indicative of multicollinearity issues because - and % tones are very often the same, i.e. L- L%. Additionally, utterances by speaker AE013 were deleted, because for that speaker no information on the interviewer was present. Also, the pragmatic functions INSEC, CHECK and EMP were conflated to a category labeled “other” because only the CzE and GerE speakers produced many instances of such functions but no speakers from the other varieties of English. The following results report on 46,970 tones, a reduced data set from the original 47,442. Thus, 472 instances (0.99% of the data) were deleted in order to create a better fit for the model. The following research question and statistical hypothesis can be stated: RQ2: *Do learners of English produce more high edge tones than NSs? Does this differ depending on REGION, position in the IP, pragmatic function, interviewer influence, topic, speaker sex, interviewer sex, age, IP length, uptalk, and quoted speech?*

H0	the distribution of the high tones of the DV HL (High vs. Low) of the NNSs equals that of the NSs, and does not differ depending on the levels of the predictors REGION, POSTION, PRAGM_confl, INTERVIEWER_ INFLUENCE, TOPIC, SEX, INT_SEX, AGE, UPTALK, and QUOTED: $R^2 = 0$.
H1	the distribution of the high tones of the DV HL (High vs. Low) of the NNSs is larger than that of the NSs, and differs depending on the levels of the predictors REGION, POSTION, PRAGM_confl, INTERVIEWER_ INFLUENCE, TOPIC, SEX, INT_SEX, AGE, UPTALK, and QUOTED: $R^2 > 0$.

The logistic regression was restricted to take the NBrE NS-norm as the intercept. “Although logistic regression is quite robust with regard to small amounts of multicollinearity, in situations of very similar operationalizations it is advisable to select one variable that is the most justified theoretically” (Levshina 2015: 273). Therefore, SPEAKING_STYLE was left out as a predictor variable because that is sufficiently covered by the pragmatic function and interviewer influence, i.e. handing over of turns (pragmatic function) and post-interactions (interviewer influence) are more common in dialogues than in monologues. AGE and IP_LENGTH were the only variables that were not significant, and were thus deleted from the final model. The vif() function revealed no issues with multicollinearity with the remaining variables. Bootstrapping of the data was performed with validate() to check for overfitting of the data. The optimism in the slopes and R^2 is less than 0.005, and overfitting is, therefore, not an issue. The final model achieved 71% accuracy, compared to a baseline of 50%. Table 55 (see Appendix) gives an overview of the results of the logistic regression. The final model included the following variables: (HL ~ REGION * POSITION + PRAGM_confl + INTERVIEWER_INFLUENCE + TOPIC + SEX * INT_SEX + UPTALK + QUOTED. Baseline 1: 0.52, Baseline 2: 0.50. Accuracy=0.71, precision=0.71, recall=0.73). The overall C-score for this model was 0.76. Hosmer and Lemeshow (2000: 162) propose that C-scores between $0.7 \leq C<0.8$ are an acceptable discrimination. Thus, the discrimination of the proposed model discriminates reasonable well. The logistic regression revealed that the odds of High (H) vs. Low (L) tones in ip-medial position (accent) in the NBrE variety of female speakers in the continuation function talking about a country (topic) in pre-interactions to a female interviewer in non-quoted and non-uptalk utterances are approximately 5.04. Thus, in this type of context, the chances of a high tone are 5.04 times greater than those for a low tone. The odds of a high vs. low tone as pitch accents in NBrE females in the context mentioned above are approximately 5.04 times higher than those in the different regional varieties of English, other variables being controlled for. The results of the logistic regression of the model are visualized in the effects plots in Figure 14.

A chisquare test was used to determine the overall significance of the predictor variables. The first plot in Figure 14 shows the different pragmatic functions ($p<0.001$***) and how the model predicts these functions to be accompanied by high tones. The listing function includes the most high tones of all functions, even though continuation and listing are quite similar. Finality is the pragmatic functions with the least instances of high tones. Utterances signaling a handing over of a turn also prefer low tones. The second plot shows that the speakers tend to use high tones more in post-interactions, i.e. in the dialogic parts, right after an interviewer turn, which indicates a generally raised f0 range, i.e. entrainment. The overall significance of the predictor INTERVIEWER_INFLUENCE is significant ($p=0.007973$**). The third plot shows the prediction of high tones by topic ($p<0.001$***). While entertainment and experience are not significantly different from those who speak about a country, the topic “self” has the least high tones of all topics. That can be explained by the fact that “self” can only appear in the dialogic part (Task 2), while country, entertainment, and experience can appear in both positions, but mostly in the monologic part (Task 1). As mentioned above, the HOT function (which is mainly used in the dialogues) is predicted to prefer low tones as compared to CON utterances. Therefore, if speakers talk about themselves in the dialogic part of the interview they may use the highlighting function of intonation (by the use of high tones) less and may signal the handing over turns by the use of low tones instead. Possibly because they may have answered the question sufficiently and wish to yield the turn. Since the topic “self” also appears later on in the recordings, this may also be a possible sign of fatigue on the part of the interviewee. Plots four and five show that high tones are more frequent in uptalk and in quoted speech (both $p<0.001$***). Plot six shows how high tones are predicted to be used as pitch accents (ip-medially) and edge tones by the different varieties of English (POSITION = $p<0.001$***). The blue line shows that MurE speakers produce significantly more high tones compared to the NBrE intercept. All other speakers use comparatively fewer high tones as pitch accents, specifically the GerE and the SAmE speakers.

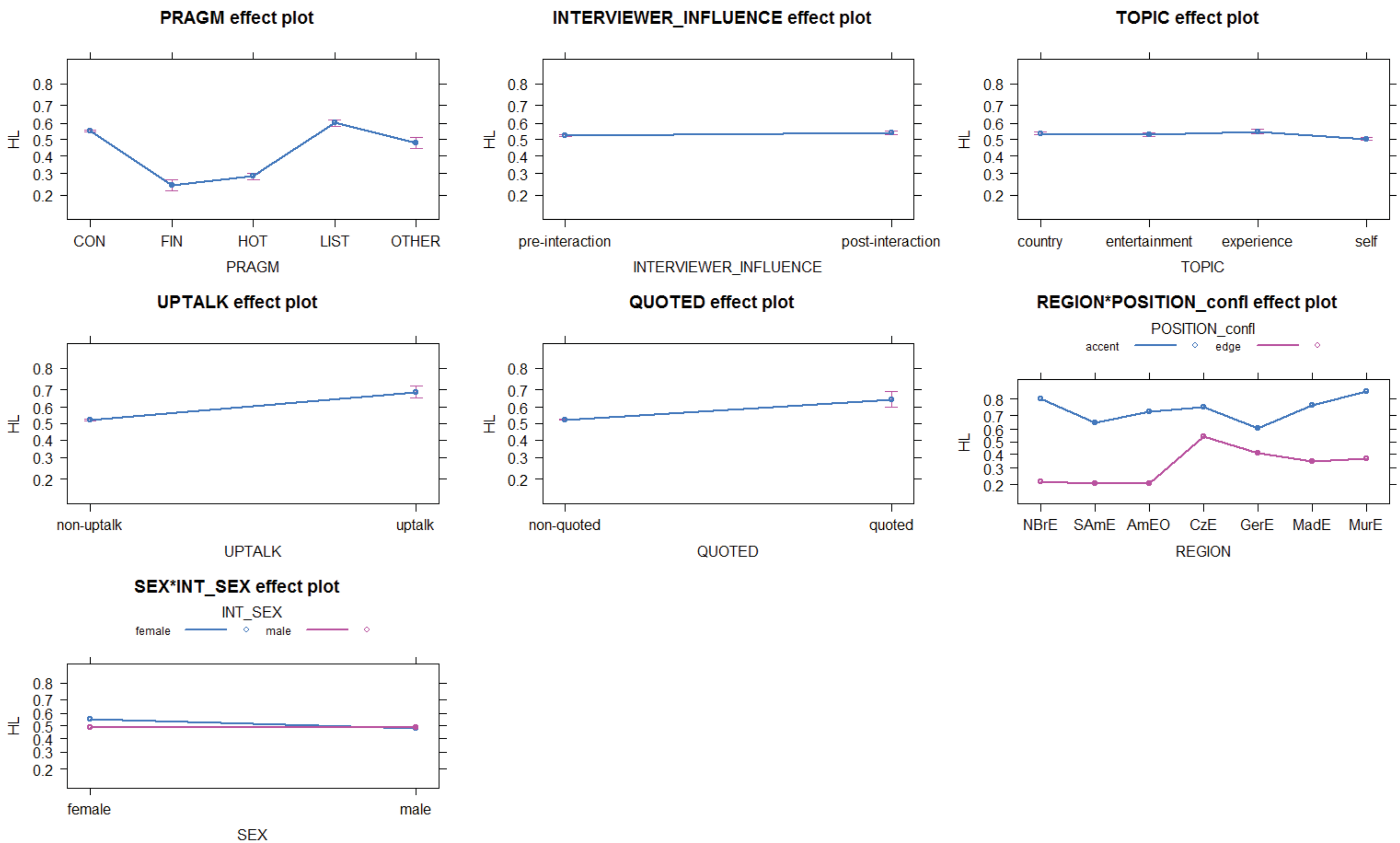


Figure 14. Effect plots for multiple logistic regression results for the distribution of high tones

As far as the edge tones are concerned, all learners produce more high tones as edge tones, specifically the CzE and GerE speakers compared to the NBrE intercept (see pink line in Plot 6). In accordance with the theory-driven expectations, H tones are more frequent in ip-medial position. H-tones are more frequently used by CzE and GerE speakers as edge tones, other factors being controlled for. The final plot (Plot 7) shows that generally there are differences between female and male speakers in the usage of high and low tones. High tones are more frequent in female speech. However, the interaction (SEX*INT_SEX, p=0.001804**) shows that when the interviewer and interviewee are both female, high tones are predicted to be used more frequently. Male speakers retain their usage of high tones in both conditions, i.e. when they speak to a male interviewer or when they are interviewed by a female interviewer. Female speakers, on the other hand, use high tones the most in female-female conversations and the least in female-male conversations. However, the confidence intervals for male-male conversations are very large, which points towards accommodation features by female speakers. In sum, H1 can be confirmed for all NNSs for RQ2, which stated that all learners will produce more high edge tones than NSs.

In order to address RQ11b: *Do learners produce fewer high edge tones after a stay abroad?*, another logistic regression was built including only the complete NNS data (CzE, GerE, and MadE) and all learner variables. The results of the predictor variables are visualized in the effects plots in Figure 15. The final model included: HL ~ REGION * POSITION + PRAGM_confl + TOPIC + SEX + UPTALK + QUOTED + STAB + YOEAS + NO_OFL (Goodness-of-fit statistic (C=0.71). Baseline 1: 0.57, Baseline 2: 0.51. Accuracy=0.67, precision=0.68, recall=0.77). The first plot in Figure 15 shows the same distribution by pragmatic functions as the first logistic regression including all speaker groups (all results for PRAGM are highly significant). For the topics entertainment and self, high tones are significantly less likely (both p<0.001***) and experience is not significantly different from those speakers talking about a country (Plot 2). Male learners use high tones significantly less (p<0.001***) (Plot 3). Uptalk (p<0.001***) and quoted speech (p=0.00197**) both lead to the production of more high tones (Plot 4 and 5). The more the learners have learned English in school, the lower their production of high tones becomes (p=0.00240**) (Plot 6). The more foreign languages the learners speak, the lower their production of high tones are predicted to be (p<0.001***) (Plot 7). Plot 8 shows that the GerE speakers are predicted to produce more high edge tones (p=0.01440*) compared to the CzE intercept, while MadE speakers are predicted to produce significantly fewer high edge tones (p<0.001***). Plot 8 also shows that GerE speakers produce high pitch accents less frequently (p<0.001***) than CzE speakers (intercept) and MadE speakers produce more (p=0.02725*). The interaction between REGION*STAB is highly significant (Plot 9). Plot 9 shows that CzE speakers with no stays abroad produce the most high tones, and GerE speakers produce fewer high tones if they have not been abroad (p=0.00782**) but there is an increase in high tones after a short stay abroad (p<0.001***). For the MadE speakers there is a decrease in the use of high tones after a short stay abroad (p<0.001***). Thus, the results for RQ11b yield contradicting or unsatisfactory results for the different learner groups. Only for the CzE learners an improvement can be observed but a contradictory development can be observed for the GerE speakers and virtually no change can be seen for the MadE speakers.

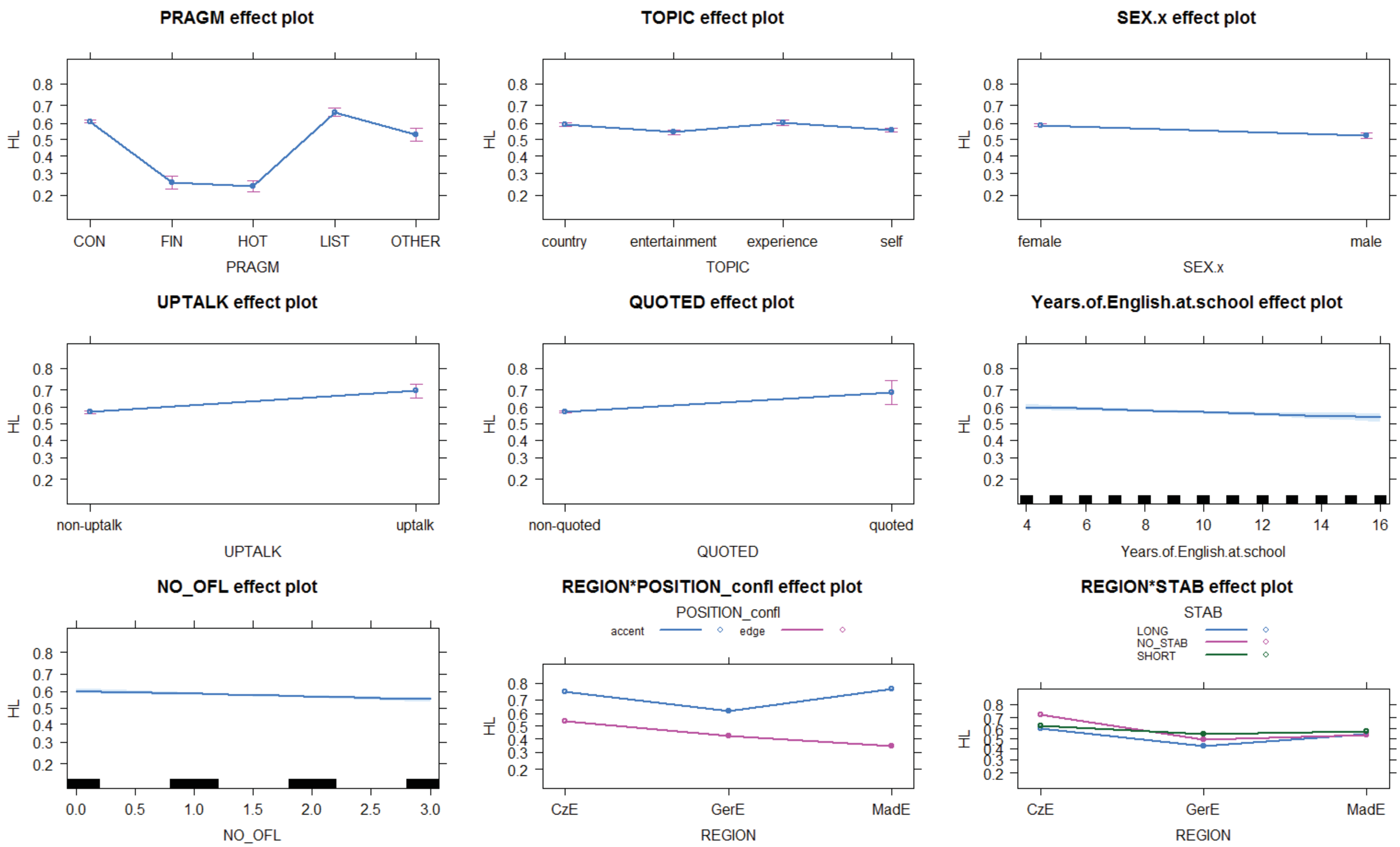


Figure 15. Effect plots for multiple logistic regression results for the distribution of high tones for the complete learner data only

4.1.3 Pitch accents

Now that the conflated tone frequencies have been determined, this section looks at the more detailed tone labels by position, i.e. pitch accents (non-final), and edge tones (final). These sections, however, only include additional descriptive statistics of the more detailed tone labels. Inferential statistical analyses have been conducted in Sections 4.1.1 and 4.1.2 with the conflated tones. Due to data sparcity issues, no reliable tests can be conducted with the detailed ToBI tone labels. Table 55 (see Appendix) gives the median frequencies of the individual tone labels in phw. As one can observe from Table 55 (see Appendix) and Figure 16, the most frequently used pitch accents are the simple H* and L* pitch accents by all varieties of English. Bitonal pitch accents are equally uncommon in all seven speaker groups. CzE speakers make a greater use of high tones in these positions and tend to produce more different phonetic variants of the H* accent, i.e. downstepped, upstepped, and high pitch accents, compared to the other speaker groups. While CzE and SpE speakers tend to produce high pitch accents more often, German learners produce them the least frequently. While NBrE and AmE speaker groups seem to exhibit very similar usages of high starred tones, the NBrE speakers differ in their use of low starred tones, which are produced quite infrequently in comparison to the SpE speakers.

In sum, the simple high and low pitch accents are the most frequent in all groups. The high pitch accent is the most common in a starred position for all groups. The Czech learners tend to produce more of all types of high pitch accents compared to the other groups (^H*, H*, !H*). Spanish learners also tend to produce H* more often than the other learners but overall they come very close to the usage of NBrE speakers for both H* and L* accents. German learners use the least high starred accents, but come very close to the usage of AmE speakers of low and high starred accents.

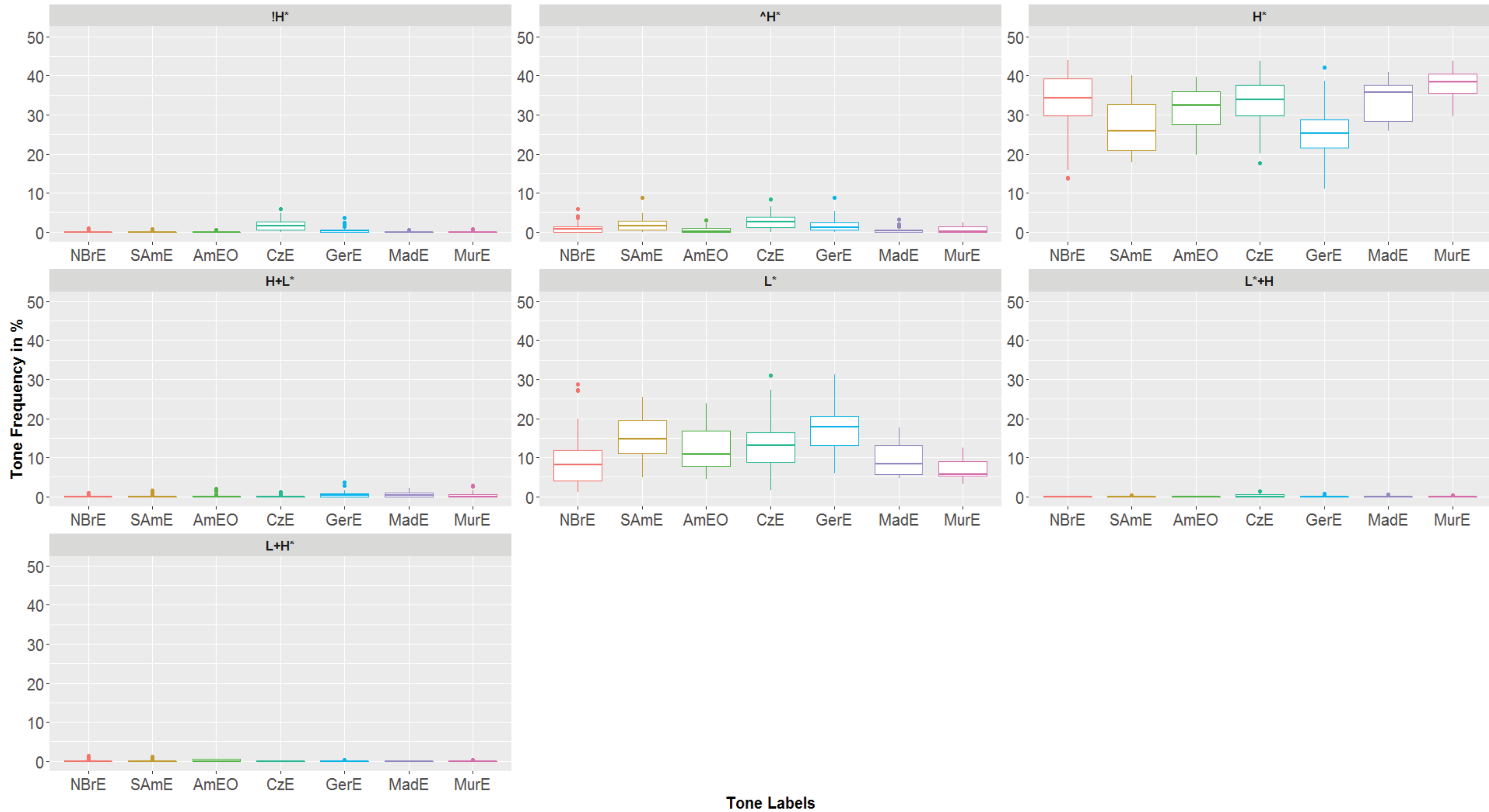


Figure 16. Boxplots for frequency of four simple and three bitonal pitch accent types (in percent) by REGION, normalized: tone frequency/number of all tones by speaker*100

4.1.4 Phrase accents

The phrase accent distributions are shown in Table 56 (see Appendix) and visualized in Figure 17. As can be observed from Figure 17, low tones are the most frequent phrase accents in all corpora, but most frequent in AmE and then NBrE speech. All learners produce low phrase accents less frequently compared to NSs, especially the CzE speakers (Plot 4). High tones seem to be slightly more popular in non-native speech sorted in order of frequency: H-, !H- and ^H-, especially with Czech learners of English.

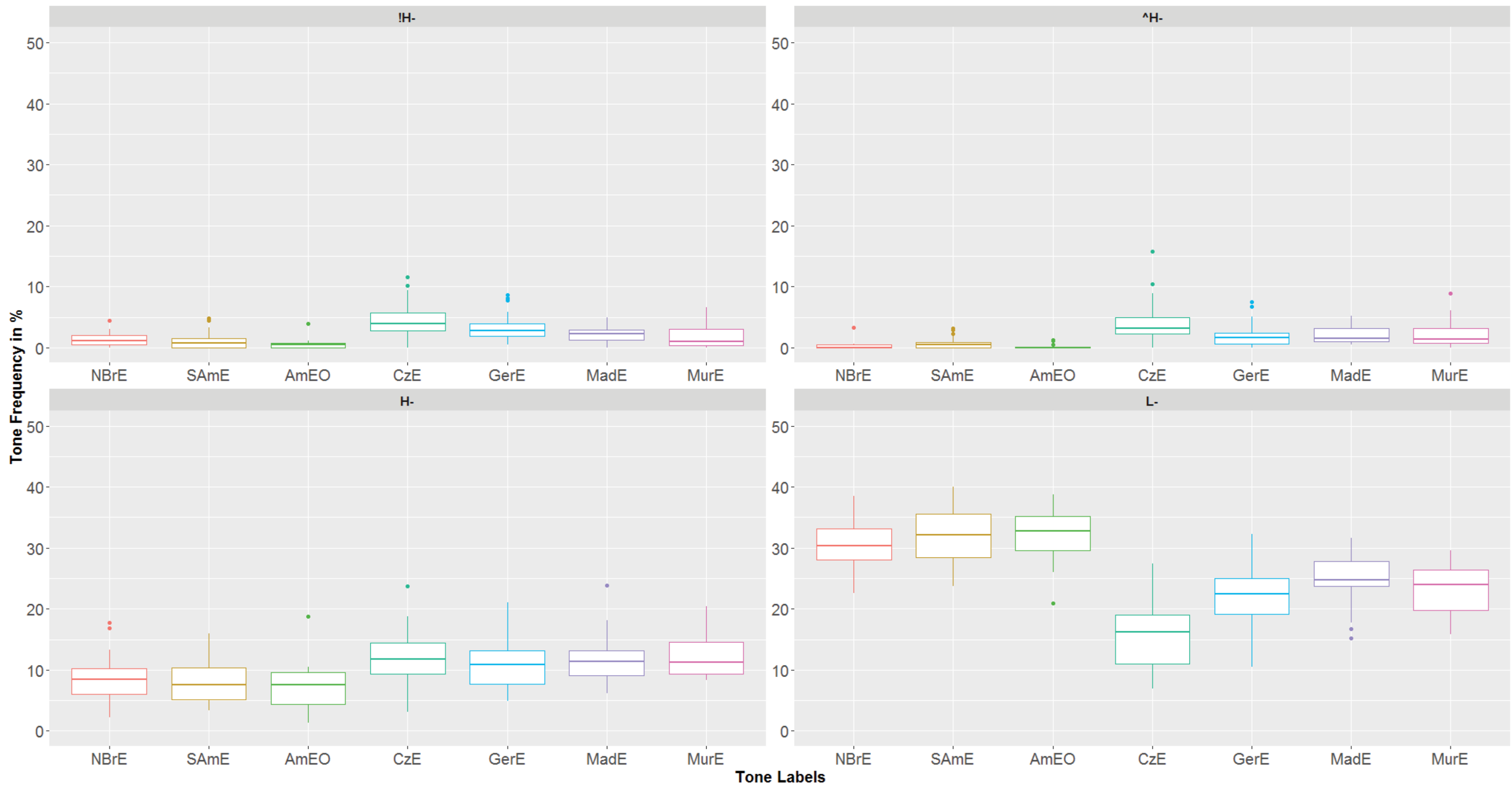


Figure 17. Boxplots for frequency of four simple phrase accent types (in percent) by REGION, normalized: tone frequency/number of all tones by speaker*100

4.1.5 Boundary tones

This section provides an overview of all initial and final boundary tones. First, the medians and IQRs of final boundary tones are given in Table 57 (see Appendix) and the results are visualized in Figure 18. The CzE speakers are the ones with the fewest instances of low boundary tones (Plot 5). The upstepped high boundary tone (^H%) is mostly produced by the NNSs in the data set, and mostly from GerE female speakers (Plot 3). While downstepped high boundary tones are relatively infrequent, they are used the most by Czech learners and also a few German learners of English (Plot 1). Initial high boundary tones are also quite infrequent but used more often by NBrE, CzE, and GerE speakers (Plot 2). While upstepped high boundary tones are nearly non-existent in native speech, they are more commonly found in non-native speech, especially in CzE and GerE (Plot 3). However, the IQRs are quite high for all values and there is considerable speaker variation. All NNSs produce high boundary tones (H%) slightly more often compared to the NSs (Plot 4). Low boundary tones are the most frequent in all corpora, although NSs use these types of tones more often compared to NNSs, with CzE speakers using them the least, and SpE and GerE speakers coming closer to the native-usage (Plot 5).

In sum, the low boundary tone is the most frequent in all corpora. However, all three learner groups use low boundary tones less frequently and instead opt for different realizations of high boundary tones. CzE and GerE speakers are very similar in their distribution of initial and final boundary tones and SpE speakers produce more native-like boundary tone frequencies.

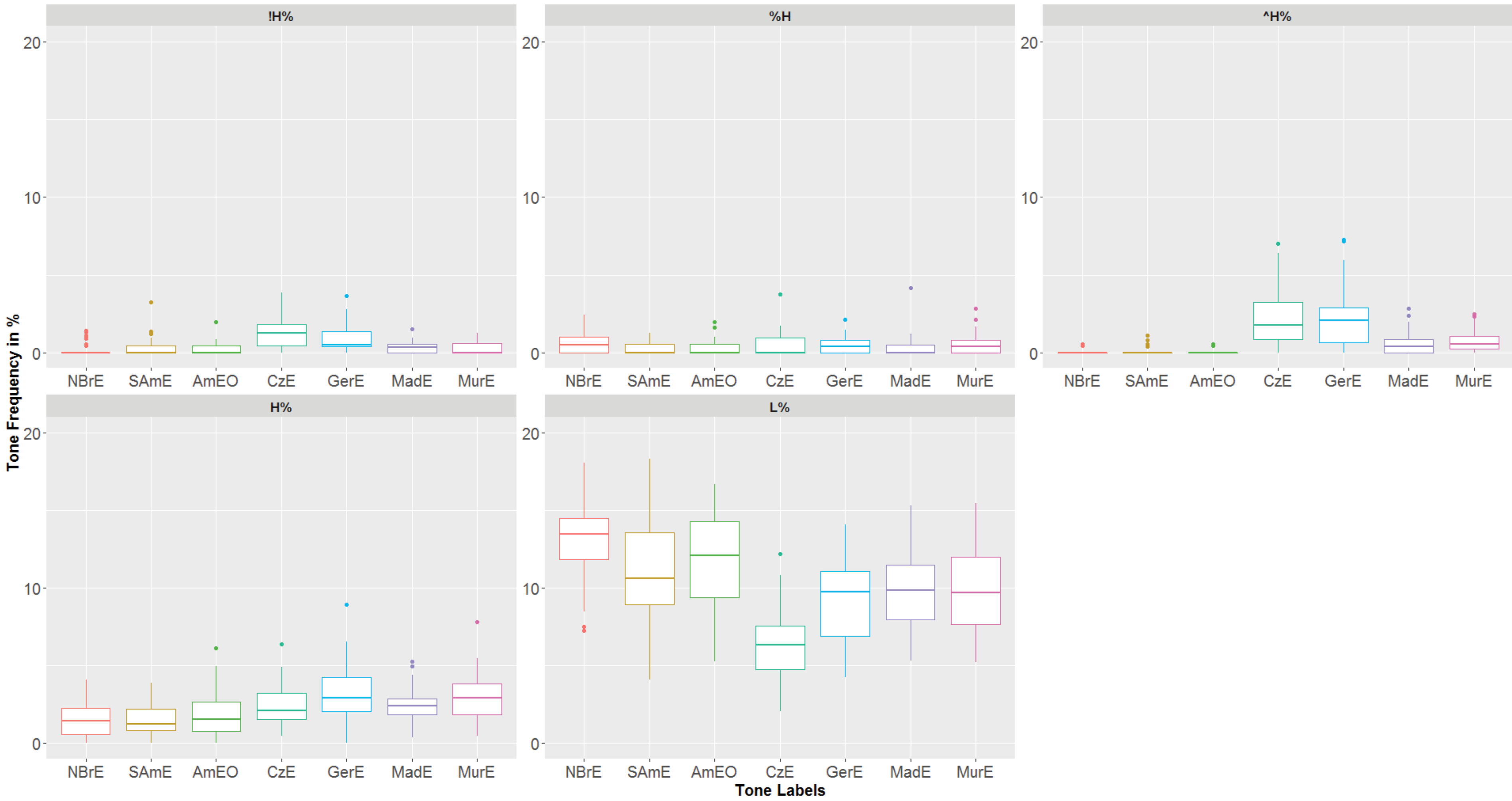


Figure 18. Boxplots for frequency of four final boundary tones and one initial boundary tone (in percent) by REGION, normalized: tone frequency/number of all tones by speaker*100

4.1.6 Tunes

In order to identify the most important factors that determine tune choice, the present section examines the distribution of tunes by ips across the sub-corpora according to pragmatic variables, in final and non-final position, and sociolinguistic variables (e.g. region and sex). Patterson (2000: 98) measured sentence-initial high pitch to final low pitch, sentence-initial high pitch to valley, non-sentence initial high pitch to final low pitch, and non-sentence initial high pitch to valley for f0 span. A similar approach has been adopted in the by-tune and by-position analysis in the present study. Even though I used the conflated tunes, i.e. HH, LL, HL, LH, LHL, and HLH, it should be kept in mind that every tune can include several H* and/or L* landmarks and each individual tone within a tune was included in the overall mean values. The present section, however, only includes descriptive statistics. Inferential statistical analyses have been conducted in Sections 4.1.1 and 4.1.2 with the conflated ToBI labels. Due to data sparcity issues for some varieties of English for the detailed tune labels, no reliable tests can be conducted with the tune patterns. Additionally, the conflated tunes will be considered more closely in the analysis of their f0 range in Section 4.3.

In total, there are 12 different tune patterns that were identified in the present data set. There are six simple tunes, i.e. high-low (HL), high-high (HH), high-flat (HF), low-low (LL), low-flat (LF), and low-high (LH). These tunes can consist of two to nine tones, but for instance in the HL-tune, the general pattern is from a high pitch accent (H*) to a low edge tone(s), i.e. low phrase accents (L-) or a low phrase accent plus a low boundary tone (L-L%), depending on the position within the ip (IP-final or IP-non-final). The opposite pattern would be the low-high tune (L* H-, L* H-H%, L* L-H%), from a low pitch accent to high edge tones. High flat/high-high and low flat/low-low are those tunes that do not change their ToBI label throughout the tune, e.g. high pitch accent and high edge tones (H* H-, H* H-H%) or low pitch accent and low edge tones (L* L-, L* L-L%). HF- and LF-tunes include ips with an f0 span of 10 Hz and under, and HH- and LL-tunes include those with 11 Hz and more. There are also six complex tunes such as HLH, LHL, LH-LH, HL-HL, HL-HL-HL, and HL-HL-H with two to seven tones per ip. In the following summary of the data, only the six simple tunes and two complex tunes (HLH, LHL) will be considered, and the other four complex tunes were dropped from the descriptive and inferential analysis (see Section 4.3) because of infrequency. In total, there are 110 instances of the complex tunes, which is 0.59% of the total data (n=18,527). Therefore, Table 58 (see Appendix) is based on the eight most commonly used tunes across the corpora, sorted by REGION and SEX in relative frequencies. The HL-pattern (a general declination trend including: H* L-, H* H- L%, and H* L- L%) is the most frequent pattern (n=8,109 out of 18,527 total ips) for ips in the entire corpus. This distribution in Table 58 (see Appendix) points towards the following trends: HL- is the most frequent tune in all varieties, native and non-native. However, generally the HL-tune is more frequent in NBrE female and AmE male speech than in non-native speech. Only the CzE speakers deviate from this pattern and they almost equally divide their ip tune usage between HH- and HL-tunes. The HH-tune is the second most frequent tune for all other females in the present data set, except the SAmE females, who seem to stick out with 'more typically male tunes'. While the females across all corpora stick to HH-, some LL- (GerE females), and LH-tunes (MadE females), most male speakers use the LF-tune second most frequently. Interestingly, while GerE and CzE speakers seem to group together in terms of tune frequencies (especially LH-tunes), MurE speakers (male and female) seem to group more with NBrE females. SpE, AmE, and NBrE speakers, including both male and female speakers, always prefer the HL-pattern over the other patterns. The HH-tune (H* H- H% or H* H-) is the second most frequent tune in the corpus. However, this stems mainly from the preference of the NNSs for such a tune, especially the CzE (both male and female) and MadE female speakers. The GerE males seem to prefer LF-tunes (L* L-L%), just like the AmEO males. The LH-tune (a general upward trend in an ip including: L* H-, L* L- H% and L* H-H%) is in general quite rare in NBrE speech, as in the male samples of all learner corpora and it is rather infrequent in general in the entire sample, compared to the flat- and HL-tunes. The SpE speakers (female and male) are a quite homogeneous group when it comes to the proportions of HL- (female: 50% vs. male: 51%), HH- (21% vs. 19%), HF- (8.88% vs. 14.7%), LL- (5% vs. 2%), LF- (5% vs. 7%), LH- (5% vs. 2 %), and the other two complex tunes. The CzE speakers also exhibit extremely similar within-group percentages of male and female speakers with HL- (female: 31% vs. male: 33%), HH- (31% vs. 24%), HF- (9% vs. 16%), LL- (7.5% vs. 9%), LF- (4.7% vs. 3.5%), LH- (9.77% vs. 7.71%), and the other two complex tunes. The GerE male and female speakers are also quite similar in the proportions of tunes but the reverse pattern for flat tunes can be observed, i.e. female speakers prefer HH-tunes (22.35%) and males LF-tunes (25.89%). While many parallels can be found between GerE and CzE speakers, the AmE and NBrE speakers show quite the opposite findings. When it comes to the HL-tune, AmE females are more similar to the NBrE males, and AmE males are more like NBrE females. In general, the NBrE female speakers exhibit little variation when it comes to their tunes within ips, i.e. 66.19% are HL-tunes, 14.01% are HH-tunes, and 6.85% are LL-tunes. While the two AmE groups behave extremely similar in male speech, there are larger differences between the two female groups. The AmEO group is closer to the NBrE females with 50% HL-tunes, 15% HH-, 9% LF-, and 8% LL-tunes. The SAmE females, on the other hand, produce fewer HL-tunes (44%), fewer HH-tunes (10.5%), more LF-tunes (17.19%), and more LL-tunes (13.18%).

If tune frequency is calculated in relative frequencies (in percent) by final POSITION, SEX, and REGION, the following picture emerges (see Figure 19; see also Table 59 in the Appendix). Of all tunes in final position, HL is by far the most frequent across all corpora. NBrE females seem to almost exclusively use the HL-pattern in final position (70.58% of all IP-final tunes). In comparison, NBrE males use the HL-tune by 20% less in IP-final position (50.92%), which is comparable to both sexes from the MurE corpus, MadE females, and both AmE groups, except the SAmE females, who group more with the GerE and CzE speakers. The groups that use the HL-pattern the least are the following: GerE (both sexes), MadE males, SAmE females, and CzE (both sexes). In sum, the HL-pattern is the most frequent in all varieties, but less so in most non-native varieties than native ones. The question that remains is which tunes the learners opt for instead of the HL-tune. The direct opposite would be the LH-tune, which is indeed more frequent in almost the same pattern as in those groups with fewer HL-frequencies. The LH-tune in final position is most common in female GerE (16.33%) and CzE females (11.95%). Interestingly, the SAmE females (7.93%) and AmEO females (7.84%) also get quite close to the CzE usage of the LH-pattern in final position. In NBrE the pattern is almost non-existent. In all corpora, the more complex HLH- and LHL-patterns are equally infrequent.

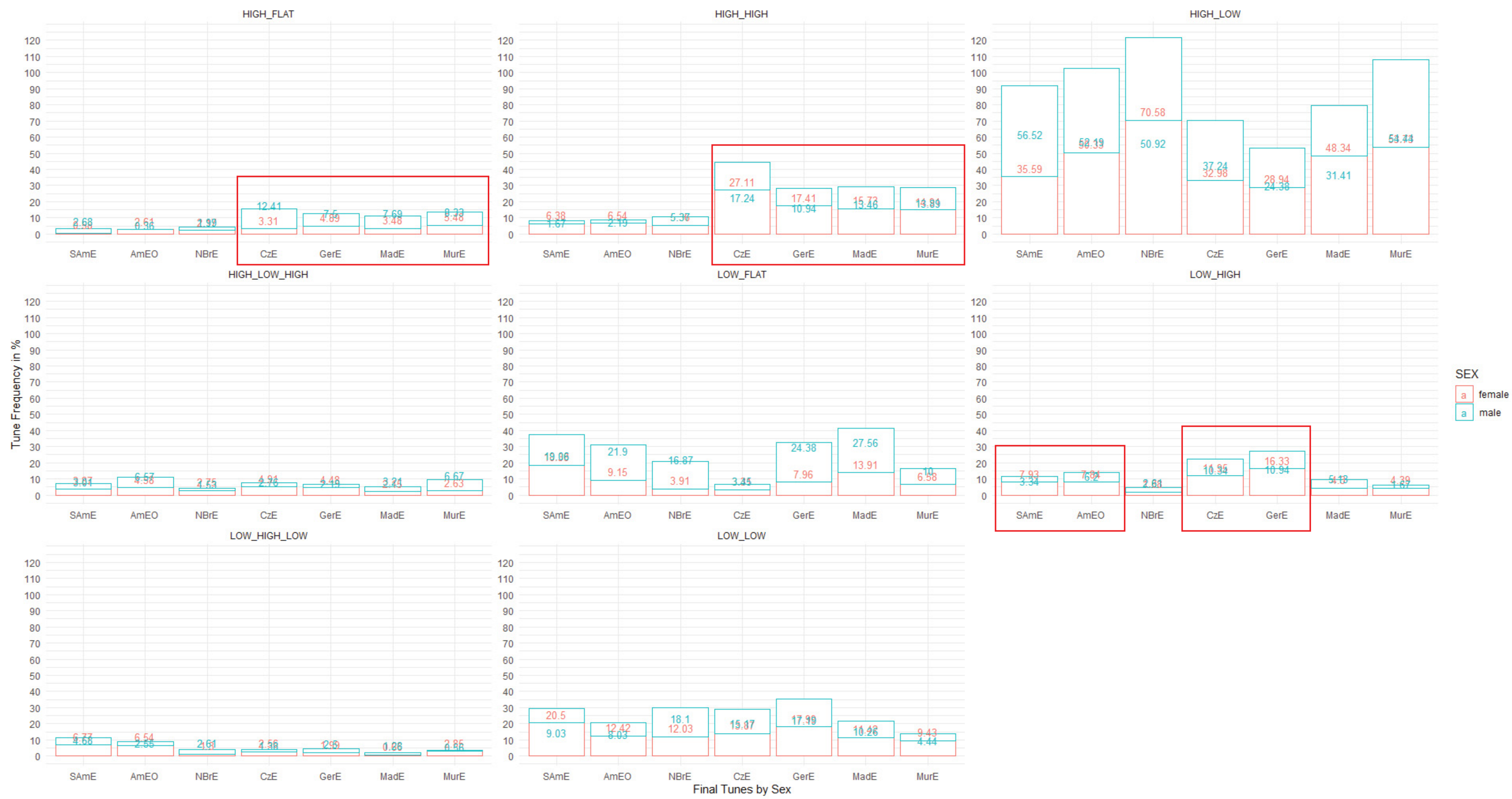


Figure 19. Barplots for tune pattern by REGION and SEX in final position (most salient features in red blocks)

Therefore, the even more complex tunes of HL_HL_HL, and the like, were excluded from the overview and further analyses, because they were even rarer. HH- and LL- are the patterns that do not change the tone label throughout an ip, i.e. they remained either in a high-register (HH or HF) or a low-register (LL or LF). Tunes of an f0 span of 11 Hz and over were considered as HH- or LL-tunes. Those tunes with an f0 span of 10 Hz and under were considered flat tunes, i.e. a large change in f0 span cannot be observed. If the HH-tunes are considered, they are more frequently used by NNSs than NSs. The same can be observed for the HF-tunes; however, these are comparatively less frequent than the HH-tunes. The CzE speakers are those with the most frequent HH-tunes (27.11% in female speech and 17.24% in male speech). CzE males also seem to be the ones with the most instances of HF-tunes (12.41%). Generally, the more flat tunes in both low- and high-registers seem to be more common in male speech than in female speech. The LF-tune is also very frequent in the MadE speakers (27.56% from male speakers and 13.91% from female speakers), and GerE males (24.38%). CzE speakers rarely use the low-flat tune, they more commonly use a LL-tune with a wider f0 span (15.17% male and 13.87% female). Generally, the LL-tune is used similarly across all corpora, albeit the most are used by GerE and the least by the MurE groups.

The tunes in non-final position by REGION and SEX in relative frequencies, depicted as bar plots in Figure 20 (see also Table 60 in the Appendix). In sum, tune trends are similar in both positions (final and non-final):

- HL is the most common tune, especially in native speech (+ Spanish).
- HH (narrower and wider f0 span) is the more common tune in non-native speech, (especially female speech).
- LL (especially narrow f0 span ones) is the more common tune in male speech and for AmE females.
- LH is very infrequent in all corpora - albeit more common in CzE and GerE (+ AmE female) especially in IP-final position.
- HLH and LHL tunes are extremely rare in both positions.

The focus now shifts to the finer distinctions between the low-high vs. high-low and distinguish between H* L-, H* H- L%, and H* L-L% (High-Low) and L* H-, L* L-H% and L* H-H% (Low-High) in final and non-final positions, as well as their distinction by pragmatic function. The high rises ending in H-H% will be called “simple high rises” and the L-H% “expanded-range rises” (cf. Flechter and Loakes 2006: 42-43). As can be inferred from Table 60 (see Appendix) and Figure 21, the most prevalent IP-final tune is H* L-L% for all varieties. The H* H-L% tune is used by the NBrE females, SAmE males, AmEO males, and the MurE speakers (both male and female). All other NSs and NNSs use this tune equally infrequently. Thus, H* L-L% is the generally preferred pattern. When it comes to IP-non-final tunes, H* L- is the preferred option for all varieties. However, the AmE, GerE, and CzE females and CzE males seem to use the L* H- pattern more than the other varieties, even though they also generally prefer the H* L- pattern. The LH-tune in its two possible realizations (L* H-H% and L* L-H%) are quite infrequently used by most varieties. However, the simple high rise (L* H-H%) seems to be generally preferred by CzE and GerE speakers by both sexes over the expanded-range rise (L* L-H%). Interestingly, the SAmE and GerE speakers seem to use the late-rise tune (L* L-H%) equally frequent, which was one of the features that stuck out during the listening and annotation phase of the recordings.

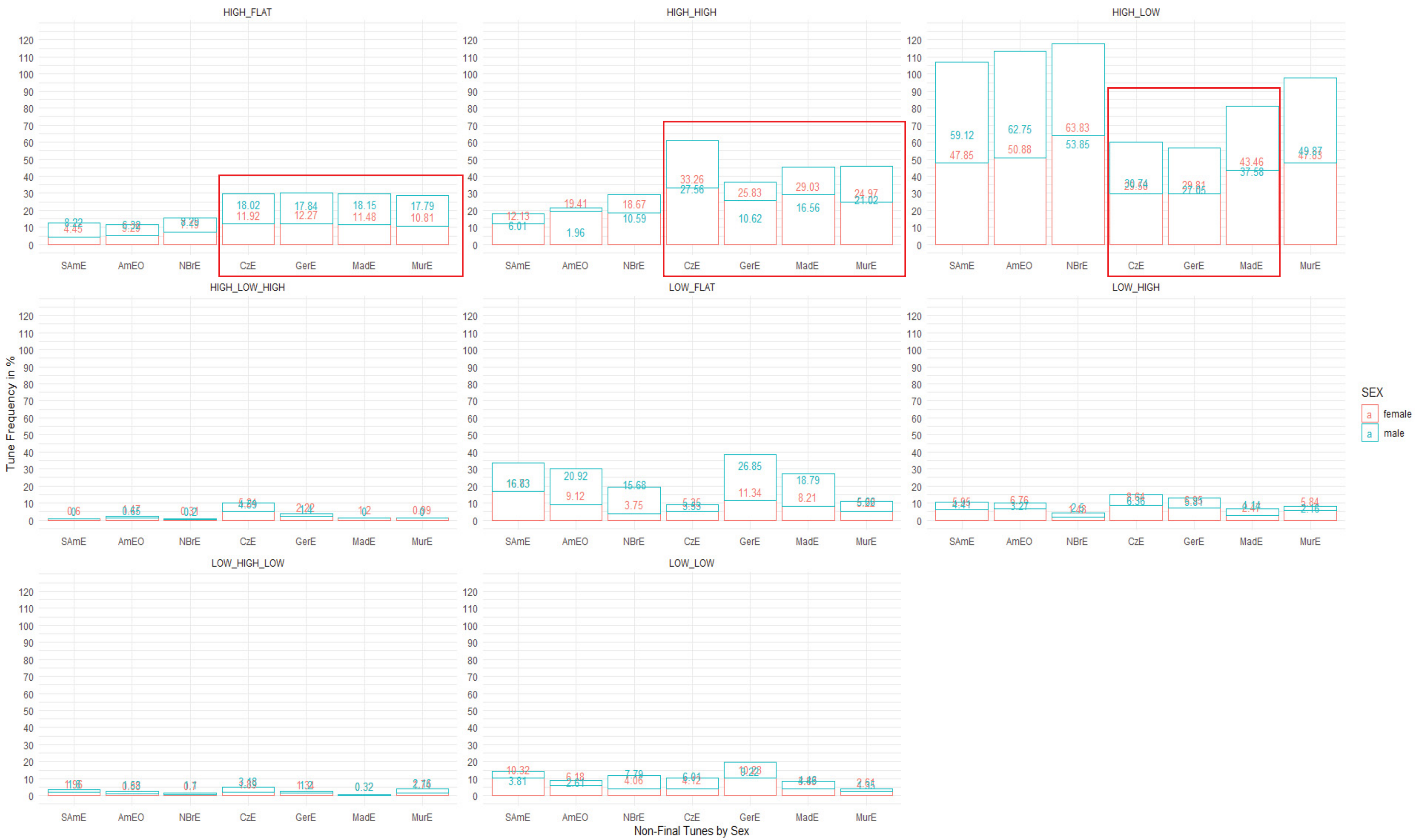


Figure 20. Barplots for tune pattern by REGION and SEX in non-final position (most salient features in red blocks)

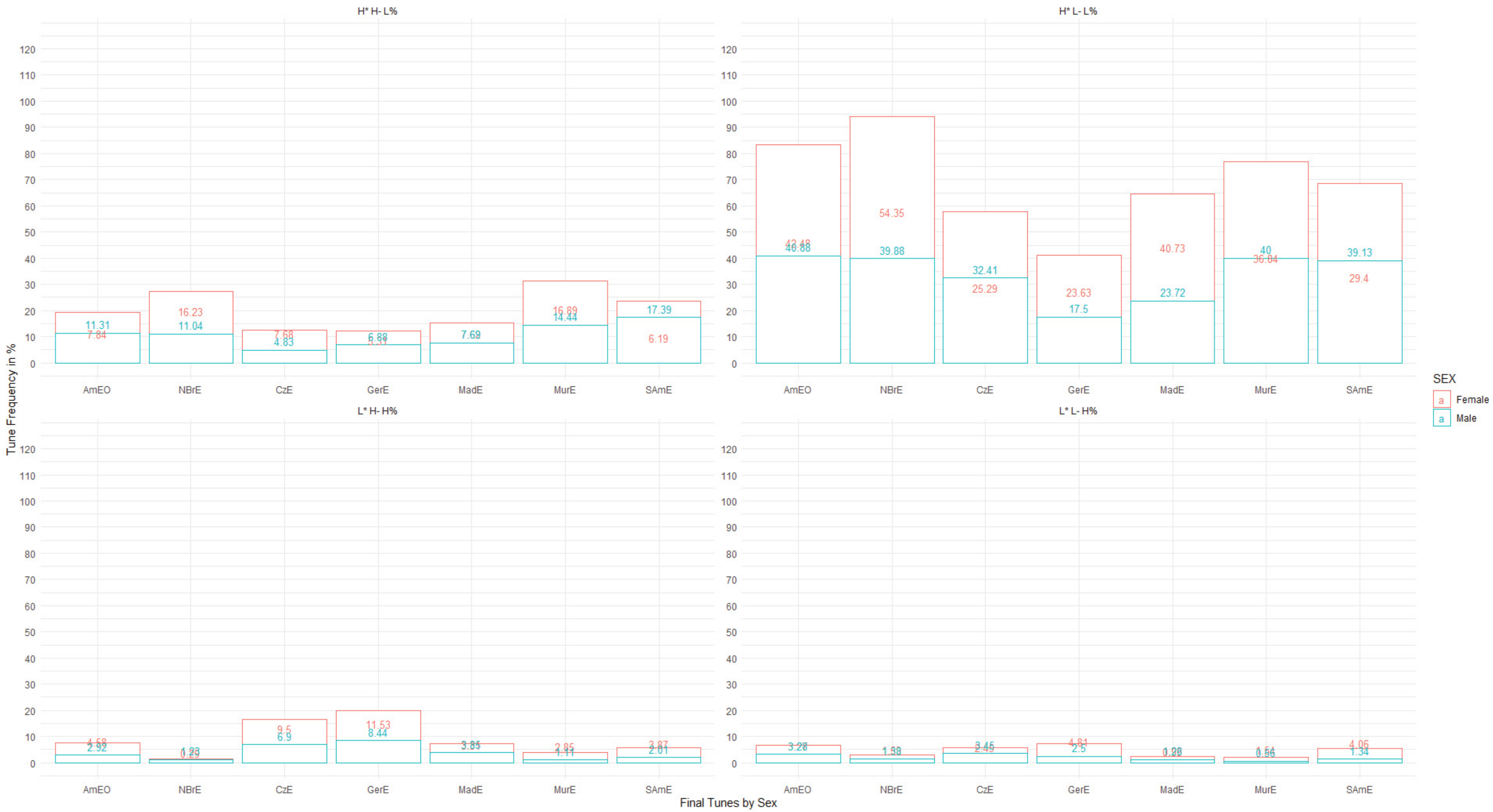


Figure 21. Barplots for final tune frequency in percent (high-low and low-high ones only) by REGION and SEX, all with the same scale. Final tunes are normalized as follows: (frequency of final tune/total number of final-tunes*100)

In order to test the proposal of whether the tunes analyzed in this study are phonologically distinct tunes, the pragmatic functions for each tune will be investigated in a second step, in order to find patterns in the distribution of the tunes. One of the few studies that include the pragmatic function of tunes demonstrated that there were no form-meaning correspondences in the analysis of HRTs and pragmatic functions in London and SoCal English speech (Barry 2008: 273). If the frequency of tunes and their pragmatic function in the entire corpus are considered (see Figure 22), it becomes clear that HL is the most common tune and the most frequent pragmatic function of that tune is to signal continuation. This was to be expected, since the selected speech samples consist mainly of continuous stretches of speech in narratives or dialogic speech. However, it is important to consider how these pragmatic functions and their corresponding tunes are distributed by the regional varieties, and speaker sex. All dialogues consist of more different pragmatic functions as compared to the monologues, with the latter primarily containing ips with a continuative function. A more detailed description of the different tunes is shown in Figure 22. Figure 22 shows the six most common simple tunes (HF, HH, HL, LH, LL, and LF), by REGION, SEX, and PRAGM (insecurity, checking, emphasis, and finality were deleted from the overview due to data sparcity).

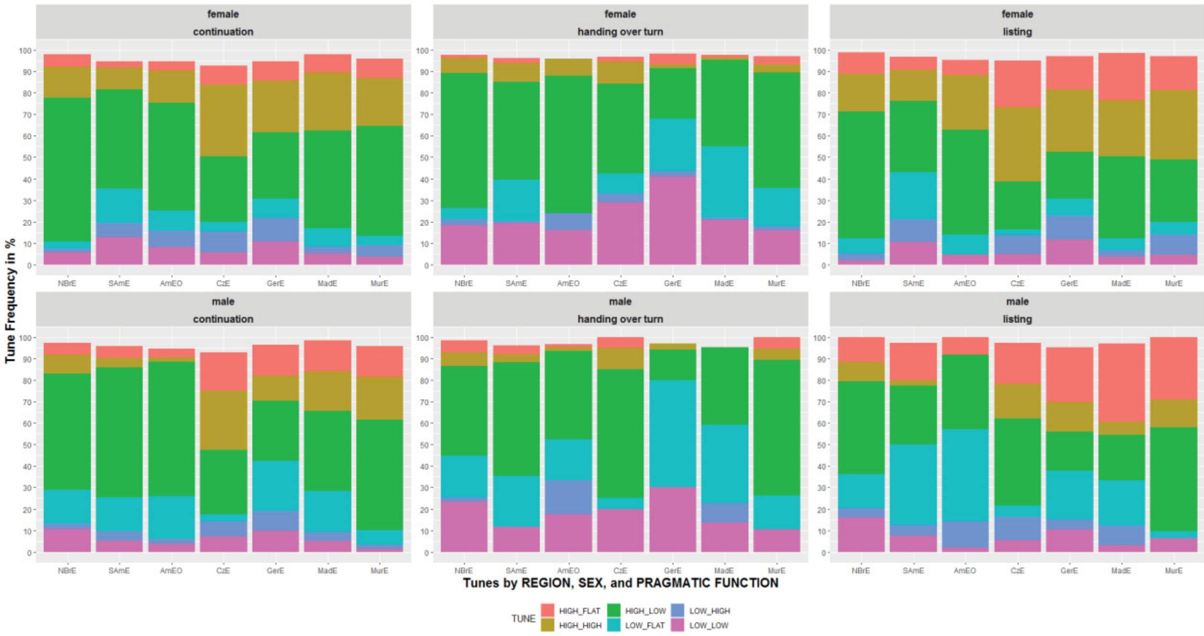


Figure 22. Barplots for tune pattern by REGION, SEX, and PRAGM

The HL-tune is by far the most common tune for most pragmatic functions across regional varieties and the two sexes. LL-tunes LF-tunes are most common in handing over turn utterances. The HH- and HF-tunes are quite rare in native speech across all functions and are very frequent in non-native speech, especially in male and female continuative utterances and in listing. The conflated tunes are analyzed from an inferential perspective in the f0 range sections.

4.2 Intonational phrasing in native and non-native speech

As has been shown in previous studies on the same learners as the ones in the present study, their fluency is a lot lower compared to the NSs of English from LOCNEC (cf. Götz 2013; Gráf 2015). Since non-native English speech has been shown to be less fluent, it therefore provides more opportunities for marking prominence (i.e. more tones) and boundaries (i.e. more IUs) than more fluent native speech. Furthermore, task differences have been observed, with learners being less fluent in all of the three LINDSEI speaking styles and NSs being less fluent in the picture description task compared to the other two tasks (cf. Götz 2013; Gráf 2015: 151). Thus, speaking style differences between monologues and dialogues, and fluency-dependent features are to be expected on IP frequency, length (in words and seconds), and speech rate. After my own classification of dialogic and monologic speech (not equal to the predefined LINDSEI sections), which is outlined in Section 3.3, 92 out of a total of 225 include mixed speaking styles (both mono and dia), 80 are pure dialogues, and 53 pure monologues. In the next two sections (4.2.1 and 4.2.2), the following hypotheses will be tested, which are all directional (one-tailed) hypotheses:

RQ3: Prosodic breaks (frequency of IUs) are expected to be more frequent in dialogues and in less fluent speech (native vs. non-native speech)

RQ4: IU speech rate: Learners will produce a slower speech rate in IUs

RQ11c: Extralinguistic factors: Learners with longer stays abroad are expected to produce fewer prosodic breaks (i.e. IUs) than those who have not been abroad

RQ11d: Extralinguistic factors: Learners with longer stays abroad are expected to produce a faster speech rate per IU than those who have not been abroad

While speaking style should explain most of the variation in the frequency, length, and speech rate of IUs, extralinguistic influences may also be at play, such as the speaker's sex, and experience with the English language (i.e. length of stay abroad). In addition to a by-speaking style and by-sex analysis, region was also analyzed as a factor. It can be assumed that the prosodic differences of the different L1s from the learners and regional varieties of the NSs may lead to differences in IU frequency, length, and speech rate.

4.2.1 IP measures

First, an overview of IP measures is given, which is followed by an ip analysis in Section 4.2.2. A summary of the descriptive statistics of IP frequency, -length (seconds, words), and speech rate is given in Table 18.

Table 18. Descriptive statistics for intonation phrases, **absolute/**normalized **(dispersion)**

Measures of central tendency	**NBrE**	**SAmE**	**AmEO**	**CzE**	**GerE**	**SpE**	Measures of dispersion in brackets
total number of IPs IPs phw	1,345 **240**	1,244 817 **146**	1,244 427 **77**	1,096 **196**	1,541 **275**	1,398 761 MadE 637 MurE **135** **129**	
median IPs/IPs phw by speakers	30/**5.4** (11/1.9)	25/**4.6** (9/1.4)	26/**4.7** (8.8/1.5)	23/**4.1** (7/1.2)	34/**6.2** (14/2.4)	30/**5.2** (6/1.1) 31/**6.4** (6.5/1.8)	**IQR**
min-max IP	17-45/ **3.1-8.1**	14-40/ **2.5-7.4**	16-39/ **2.8-7.2**	13-38/ **2.4-6.9**	16-51/ **2.8-9.1**	20-46/ **3.4-8.3** 20-51/ **4.1-9.3**	
median length (s)/WPS	4.31/**3.8** (3.7/1.3)	7.14/**2.68** (5.8/0.86)	7.12/**2.71** (6.2/0.83)	7.31/**2.66** (6.2/0.84)	5.02/**2.9** (4.5/1)	5.9/**2.79** (5.8/0.99) 5.6/**2.47** (4.8/1)	**IQR**
min-max length (s)/WPS	0.64-23/ **1.48-7.78**	1.02-44.31/ **1.23-5.32**	0.5-44.46/ **0.94-5.96**	0.9-44.5/ **1.23-5.32**	0.5-42/ **0.78-8.88**	0.8-26/ **1.11-5.8** 0.62-45/ **0.95-7.73**	
median words/WPS	16/**3.8** (13/1.3)	19/**2.68** (16/0.86)	19/**2.71** (17/0.83)	20/**2.66** (17/0.84)	14/**2.9** (12/1)	16/**2.79** (14/0.99) 13/**2.47** (9/1)	**IQR**
min-max words/ WPS	2-101/ **1.48-7.78**	2-123/ **1.23-5.32**	1-100/ **0.94-5.96**	2-123/ **1.23-5.32**	1-116/ **0.78-8.88**	3-68/ **1.1-5.8** 3-77/ **0.95-7.73**	
median of ips per IP	2 (1)	3 (2)	3 (2)	3 (2)	2 (2)	3 (2) 2 (1)	**IQR**
min-max ips	1-14	1-13	1-11	1-13	1-12	1-10 1-14	
median of tones per IP	6 (3)	7 (5)	7 (5)	7 (5)	5 (4)	7 (4) 7 (4)	**IQR**
min-max tones	3-29	3-32	3-30	3-32	3-29	3-22 3-33	

Before any more statistics are given on IP speech rate, it is important to consider previous research on the intersections of fluency, proficiency, and prosodic phrasing and the fluency and proficiency levels of the learners in the present study. Previous research has suggested that learners have a lower lexical diversity, which is caused, among other factors by the tendency of learners to avoid difficult words and phrases (e.g. Dagut and Laufer 1985; Liao and Fukuya 2004) and often use the same and simpler constructions. According to Götz (2013: 65), non-native-like lexical choices can negatively influence other fluencemes and lead to more lexical errors. Thus, considering previous research, the lexical repertoire used in the student utterances in the present study might give an indication of their proficiency levels, which in turn might reflect on their fluency, and ultimately their intonational phrasing performance. All of these levels are intertwined and may indicate a learner's proficiency level. For instance, the more a learner repeats simple words and constructions, e.g. by avoiding more complex ones, more words are needed to circumscribe, and the more a learner might pause to plan the next utterances ahead. This affects fluency and the frequency and length of IUs. IPs are the domain to which all other prosodic features (e.g. tones and tunes) are confined. Since fluency measures (filled and silent pauses, as well as elongations) have been shown to have a consistent effect on intonational phrasing in both native and non-native speech, more so than gender and speaking style differences, the interdependence on fluency might give further indications about the proficiency of the learners in the present data set. Previous research has also shown that lexical diversity improves with a stay abroad (Tavakoli and Foster 2008; Foster and Tavakoli 2009). In the present study, none of the "advancedness"-variables tested showed a strong correlation with the CEFR levels (All CEFR ratings were taken from the phonological control section) of the LINDSEI-CZ and LINDSEI-GE sub-corpora (see Table 19). Only words per second (WPS) per IP and MLR measures show a moderate positive correlation with the CEFR levels. All other correlations are weak, and do not reflect on the CEFR levels of the CzE and GerE speakers. Since the IP speech rate appears to be the measure with the best indication of a higher CEFR level, these will be tested in the present study. Additional fluency measures were also correlated with the IU variables taken from the respective 560-word samples. Since proficiency could not be accounted for all learners in the present data, correlations were tested for the known proficiency levels of the LINDSEI-CZ and LINDSEI-GE sub-corpora, with fluency measures from previous research on the same learners, lexical complexity measures of the samples, IUs, and tone frequencies, as well as months in an English-speaking country and years of English at school and university. Table 19 shows the results of a correlation analysis of fluency, intonation, and other advancedness features with the CEFR levels. Generally, the higher the CEFR level, the faster the IPs are produced (WPS). Thus the more proficient speakers can speak faster and fit more words into an IP. However, it has to be kept in mind that there are only two C2 speakers, and thus the correlations are not reliable for this category.

Table 19. Correlations between CEFR levels in LINDSEI-CZ and LINDSEI-GE and fluency measures, lexical complexity measures, IP, ip, and tone frequencies, months in English-speaking country, and years of English at school and university

Profi-ciency CEFR levels	**WPS per IP/ip**	**MLR Task_1/ Task 2** (based on CzE n=25)	**Guiraud Index (lexical complex-ity)**	**TTR (lexical complex-ity)**	**Years of English at university /school**	**IP freq.**	**Months in English speak-ing country**	**Tone freq.**	**ip freq.**	**Age**
Pearson's r CzE (n=45)	.63/.62	.46/ .40	.35	.33	.10/ -.09	-.27	-.14	-.09	-.03	0.11
Pearson's r GerE (n=19)	.31/.23	NA	.18	.22	-.11/-.03	-.08	.21	-.09	-.10	- .018
Measure	Words per IP/ip / seconds	sum of all lengths of runs / number of runs)	Types/ SQRT of Tokens	Types/ Tokens * 100		IPs phw		Tones phw	ips phw	

Even though the GerE sample (n=19) is based on only half as many speakers as the CzE sample (n=45), and it only has the CEFR levels of B2 and C1, strikingly similar correlations have been found. None of them are strong correlations, but WPS seem to be the best indicators of a higher CEFR level, but with only a moderate correlation.

Since fluency seems to be a good indicator of proficiency level of learners, further fluency variables were included in the analysis of IUs. The present study did not distinguish between different types of filled pauses *er*, *um*, and *mm*, since previous research has found only minor differences between these types of filled pauses in how the speakers are perceived in terms of certainty together with the prosody they are using (e.g. see Smith and Clark 1993 for an opposite result; perception study: Brennan and Williams 1995). Therefore, the filled pauses *er*, *um*, and *mm* were added together into the variable (FP). The silent pauses are called UP (unfilled pauses), and they include both silent (up) and breathing pauses (b). Elongations were labeled (EL). It is apparent from Figures 57-60 (in the Appendix) that most learners generally pause more often (filled pauses) and make use of more elongations and hesitations (HP) than NSs. The CzE speakers stick out when it comes to the frequency of filled pauses and elongations, and the two SpE groups (MurE speakers more so than MadE speakers) use slightly more silent pauses and hesitations. Whether these measures correlate with IP frequency and IP speech rate will be determined in the respective sub-sections. The fluency (FP_NORM, UP_NORM, EL_NORM, and HP_NORM) measures, the IP speech rate measure (WPS), and the main independent variables REGION, SEX, and SPEAKING_STYLE (and more) are tested against IP frequency (dependent variable) in the following sub-sections.

4.2.1.1 IP frequency

This section contains the analysis of the larger IUs, i.e. IPs. In total there are 6,624 IPs, with 2,665 IPs coming from the monologues and 3,959 IPs coming from the dialogues, in the entire corpus including all seven corpora and seven different varieties of English. In total 4,510 IPs come from female speakers and 2,114 from male speakers. Basic descriptive statistics for the IPs in both absolute and normalized values are given in Table 18 in Section 4.2.1, broken down by REGION. Since the corpora in this study vary by their number of IPs, IP frequency was normalized to phw. A visual inspection with the help of histograms showed that IP frequency is skewed to the right. In order to remove some of the skewness, the frequencies of all numeric dependent and some of the independent variables were logarithmically transformed (log2).

In order to answer hypothesis three, it is necessary to look at IP frequency (phw) in the seven sub-corpora. Figure 61 (see Appendix) demonstrates the descriptive results. We observe from Figure 61 that MurE and the GerE speakers produce the most IPs, and the CzE speakers produce the lowest number of IPs compared to the NSs. The MadE speakers behave similarly to the NBrE speakers, while the AmE speakers (both groups) come closer to the CzE speakers with fewer IPs phw. These differences might be due to the fluency of speakers, interview styles, speaking style differences, or a combination of these factors. This result does not only reflect native vs. non-native differences in IP frequency, especially because the MadE speakers seem to be quite native-like with regard to IP frequency, but rather differences in speaking styles, contextual factors, and regional dialect, as well as proficiency in the case of the learners. The range of IP frequency by NE regional dialect is between four and five IPs phw (two IPs for logged values). Overall there are no significant differences between the NSs from all regions (NBrE vs. SAmE vs. AmEO). In sum, IP frequencies seem to be quite similar in all varieties of English (native and non-native), especially if they are normalized and logged. Therefore, in order to analyze the results from an inferential perspective, a linear regression was fit to answer the following research question: RQ3: *Do learners of English produce more prosodic breaks (IPs) than NSs of English? Does this differ depending on the speaking style, and fluency (measures in WPS, number of silent and filled pauses, number of hesitations)?*

H0	the distribution of the DV IP frequency (IP_PHW) of the NNSs equals that of the NSs, and does not differ depending on the levels of the IVs SPEAKING_STYLE, SEX, UP, FP, HP, and WPS: $R^2 = 0$.
H1	the distribution of the DV IP frequency (IP_PHW) of the NNSs is larger than that of the NSs, and differs depending on the levels of the IVs SPEAKING_STYLE, SEX, UP, FP, HP, and WPS; $R^2 > 0$.

A linear regression was selected for this research question because there was only one data point per speaker, i.e. the specific number of IPs per speaker. In order to test RQ3, and to test whether speaking style and fluency differences have a significant effect on the dependent variable IP_PHW, a linear regression was fit. Log2(IP_PHW) was entered into the model (adjusted $R^2=0.52$, $F=15.27$ on 17 and 207 DF, $p<0.001$)[74] as the dependent variable and REGION (NBrE, SAmE, AmEO, CzE, GerE, MadE, and MurE), SEX (female vs. male), fluency measures (FP_NORM, UP_NORM, HP_NORM, WPS_IP), and the SPEAKING_STYLE (both, dia, mono) were entered into the model as the IVs. Using backward model selection, I arrived at the minimal model shown in Table 20. Model diagnostics with the help of Q-Q plots of the residuals and histograms showed no deviations from normality and homoscedasticity. Also, the corrected and uncorrected VIF-scores are small (under five), and thus multicollinearity is not a problem. A durbinWatsonTest determined no autocorrelation between the residuals with a p-value of 0.924 and a D-W statistic of 2.08. The results can be seen in Table 20.

[74] The adjusted *R*2 was used because this is standard procedure in linear regression analysis (Carver and Nash 2011: 201).

Table 20. Results for linear regression with IPs phw. Final model: log2(IP_PHW) ~ 1 + REGION * SEX + log2(UP_NORM) + SPEAKING_STYLE + log2(WPS_IP). Speech rate measure: words per seconds (WPS)

Predictors	Estimate	Std. Error	t-value	p-value	
(Intercept)	-0.70397	0.41507	-1.696	0.091382	.
REGION_SAmE	-0.18746	0.09	-2.083	0.038485	*
REGION_AmEO	-0.16048	0.12708	-1.263	0.208059	
REGION_CzE	-0.23679	0.09209	-2.571	0.010838	*
REGION_GerE	0.21425	0.08645	2.478	0.014003	*
REGION_MadE	-0.13091	0.09782	-1.338	0.182283	
REGION_MurE	0.03214	0.12019	0.267	0.789439	
SEX_male	0.06626	0.08472	0.782	0.435039	
log2(UP_NORM)	0.6381	0.06543	9.753	<0.001	***
SPEAKING_STYLE_both	0.14585	0.05672	2.571	0.010829	*
SPEAKING_STYLE_dia	0.18175	0.05301	3.428	<0.001	***
log2(WPS_IP)	0.29431	0.12467	2.361	0.019171	*
REGION_SAmE:SEX_male	0.2058	0.13881	1.483	0.139721	
REGION_AmEO:SEX_male	0.08257	0.16641	0.496	0.620272	
REGION_CzE:SEX_male	-0.04164	0.15075	-0.276	0.782666	
REGION_GerE:SEX_male	-0.23848	0.12483	-1.91	0.057461	.
REGION_MadE:SEX_male	0.10755	0.16305	0.66	0.510243	
REGION_MurE:SEX_male	-0.35108	0.1606	-2.186	0.029933	*

According to the results of the linear regression (see Table 20), most regional dialects (except GerE and MurE females) are predicted to produce fewer IPs compared to the NBrE intercept, but most of the differences are not significant (only for GerE, SAmE, and CzE female speakers, and MurE males). GerE female speakers (p=0.014003*) are predicted to use more IPs phw than the NBrE speakers, if the number of silent pauses (UP), speaking style, and WPS are taken into account. The SAmE (p=0.038485*) and CzE female speakers (p=0.010838*), and MurE male speakers (p=0.029933*) produce significantly fewer IPs phw than the NBrE intercept. However, the confidence intervals for the male speakers are very large, due to the lower number of male speakers and the results are therefore not reliable. The model also shows that the more silent pauses a speaker produces (UP) (p<0.001***), the more dialogic the speaking style is (dia=p<0.001***), the faster the IPs are produced in WPS (p=0.019171*), the more IPs are produced. Even though at first glance an R^2 of 0.52 seems to be quite low, IP frequency is an extremely variable feature, which may depend on many different factors, which could not be tested in the present study, such as personality of the speaker, experience in public speaking, different contextual factors, and/or the mood of the speaker, for instance. It is, thus, a quite satisfactory value, since it predicts that there are variational (interaction between REGION * SEX p=0.01008*), speaking style (p=0.00265**), fluency (UP p<0.001, and WPS_IP p=0.01917*), and sex differences that determine IP frequency. The results are visualized in the effect plots in Figure 23.

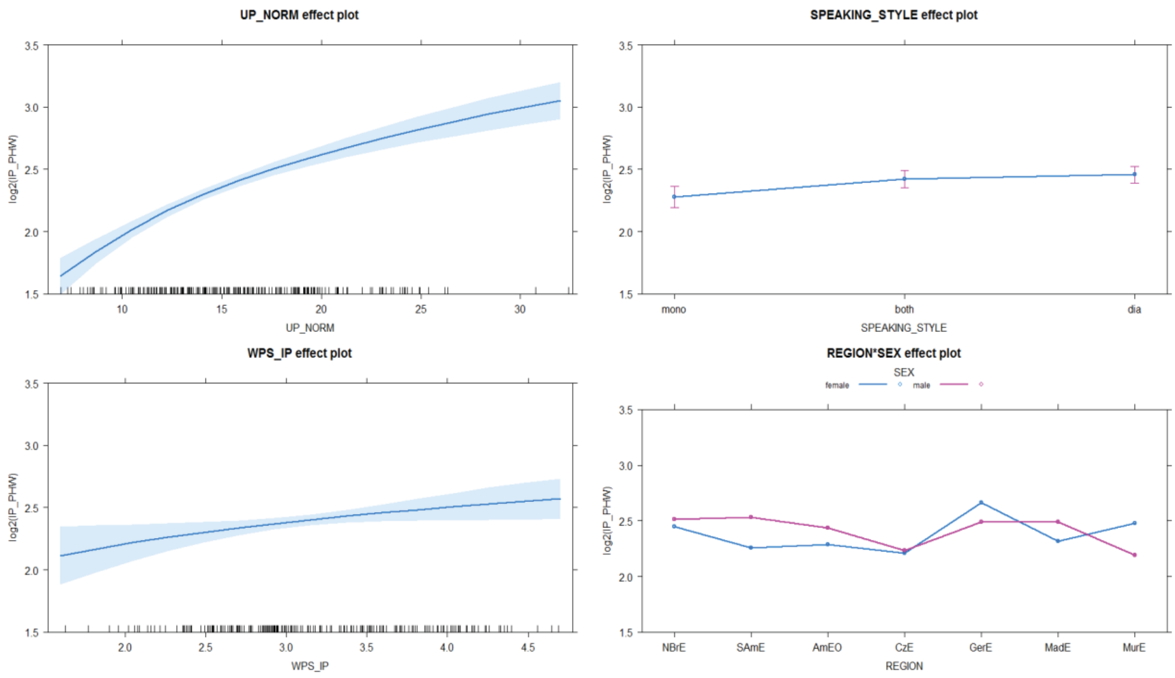


Figure 23. Effect plots for linear regression results of IPs phw

Even though the above model only showed that GerE female speakers are predicted to produce significantly more IPs in comparison to the NBrE intercept, another model was fit to answer RQ11c: *Do IP breaks become less frequent after a stay abroad?* None of the learner variables and any of their interactions were significantly contributing to the predictive power of the model (adjusted R^2=0.59, F=11.86 on 15 and 99 DF, p<0.001). However, what the second analysis shows again is that the GerE female speakers are predicted to produce significantly more IPs as compared to the CzE intercept (p<0.001***). To conclude, the linear regression analyses lead to the partial confirmation of H1 of RQ3, i.e. that NNSs are predicted to produce more IPs phw. This was only true for GerE females. The IP frequency for CzE (female), SAmE (female), and MurE (males) speakers was determined to be less than in the productions of NBrE speakers. All other differences between native and NNSs were not significant.

4.2.1.2 IP speech rate in WPS

The present section contains the IP speech rate analysis with normalized values in words per second (WPS). The measure of IP length in words was normalized by dividing the number of words of every IP by the number of words of the respective IP. This measure was moderately and positively correlated with the CEFR levels of the Czech and German learner groups and was deemed to be a more reliable measure than IP length in words and seconds because the corpus samples were of varying lengths. If normalized (length in words of each IP/no of words of each respective IP) and logged values are considered, the NSs produce faster IPs, while all learners, especially those speakers from Murcia, produce the slowest IPs (see Figure 62 in the Appendix). Even IPs in dialogues are produced faster by most regional varieties in the present study, in both speaking styles (dialogues and monologues) the NSs produce the fastest IPs in WPS compared to the NNSs (see Figure 63 in the Appendix). MurE and CzE speakers are the ones with the slowest IPs in WPS. These descriptive results are complemented by inferential statistics, in order to answer the following research question: RQ4: *Do learners of English produce slower speech rates in IPs compared to NSs? Does this differ depending on speaker sex, age, speaking style, and fluency?*

H0 the IP speech rate (WPS) of the NNSs is equal to that of the NSs, and does not differ depending on the levels of the IVs REGION, SEX, AGE, SPEAKING_STYLE, UP, FP, EL, and HP: $R^2 = 0$.

H1 the IP speech rate (WPS) of the NNSs is smaller than that of the NSs, and differs depending on the levels of the IVs REGION, SEX, AGE, SPEAKING_STYLE, UP, FP, EL, and HP: $R^2 > 0$.

Mixed models were run to test the predictive power of REGION, SEX, and SPEAKING_STYLE on IP speech rate in WPS, using the lmer function from the lme4 package (Bates et al. 2015) (see Table 21). SPEAKER was added as a random intercept. Main effects of REGION, SEX, SPEAKING_STYLE, UP, FP, HP, and EL were entered into a model (conditional $R^{2=}0.37$) without an interaction term. Adding random slopes for REGION and SEX resulted in the model's failure to converge, and were thus removed. P-values were obtained with the lmerTest package and the conditional R^2 with r.squaredGLMM(). An overall p-value for the model was obtained by likelihood ratio tests of the full model with the fixed effect in question compared to a model without the fixed effect. Further, a visual inspection of the data with residual plots did not reveal any deviations from homoscedasticity but a slight deviation from normality. Therefore, a mixed model with logged (log2) values was fit.

Table 21. Results for mixed effects model with WPS per IP. Final model: log2(WPS_IP) ~ REGION + SEX + AGE + SPEAKING_STYLE + EL_NORM + FP_NORM + UP_NORM + HP_NORM + (1 | SPKR). Speech rate measure: words per seconds (WPS)

Fixed effects:	**Estimate**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	2.39267	0.07503	31.89	<0.001	***
REGION_SAmE	-0.01129	0.02891	-0.39	0.69657	
REGION_AmEO	0.01041	0.03562	0.29	0.77035	
REGION_CzE	-0.18829	0.03215	-5.86	<0.001	***
REGION_GerE	-0.16783	0.02823	-5.94	<0.001	***
REGION_MadE	-0.23406	0.0331	-7.07	<0.001	***
REGION_MurE	-0.16817	0.04182	-4.02	<0.001	***
SEX_male	-0.01737	0.01799	-0.97	0.33563	
AGE	-0.00483	0.00293	-1.65	0.10031	
SPEAKING_STYLE_dia	0.03413	0.01198	2.85	0.00444	**
EL_NORM	-0.02165	0.00375	-5.78	<0.001	***
FP_NORM	-0.01541	0.00335	-4.6	<0.001	***
UP_NORM	-0.02089	0.00224	-9.32	<0.001	***
HP_NORM	-0.01652	0.00443	-3.73	<0.001	***

Number of obs.=6,589, Random intercepts: Speaker SD (0.0925), Residual SD (0.3511); Log likelihood f0.model =-2569, Conditional $R^{2=}0.37$.

The results are visualized in the effect plots in Figure 24. The NBrE female intercept in monologues produce among the fastest speech rate (2.39 WPS per IU logged values), and all other regional varieties produce slower speech rates per IP (except for AmEO who are slightly faster than the NBrE intercept) (Plot 1 in Figure 24). However, the differences are only highly significant for the NNSs. The MadE speakers produce the slowest IPs, then the CzE speakers, followed by the GerE and MurE speakers, all other factors being accounted for. The predictor REGION in the model is highly significant ($X^2(6)=80.3$, $p<0.001$***). The independent variables SEX and AGE do not play a significant role in IP speech rate measured in WPS (logged), although a tendency for speech rate to be slower per IP can be observed for both variables (Plots 2 and 3). Dialogues include significantly faster IPs in WPS (0.034, $p=0.00444$**) (Plot 4). All fluency measures lead to the production of slower IPs, i.e. the more elongations, hesitations, filled and silent pauses are used, the slower the IPs are produced ($p<0.001$*** for all fluency measures) (Plots 5-8). Thus, regional variation, speaking style, and the various fluency measures explain 37% of the variation in the speech rate variable per IP. Based on these results, I decided in favor of H1, i.e. the alternative hypothesis that stated that all NNSs would exhibit a slower speech rate within IPs measured in WPS.

In order to see how the learners (complete data only: CzE, GerE, and MadE) compare to each other and to answer RQ11d: *Do learners produce a faster speech rate (WPS) per IP after a stay abroad?*, a separate model was fit with the following variables: REGION, SEX, SPEAKING_STYLE, AGE, fluency variables (EL, FP, HP, and UP), STAB (long, short, no stay abroad), YOEAS and YOEAU, and number of other foreign languages the learners speak (NO_OFL). Overall, the model only explains 17% (conditional R^2) of the variation. As shown in Figure 25, all the fluency variables are the most significant as in the model above (EL and UP both $p<0.001$***). Interestingly, however, the higher the number of other foreign languages (NO_OFL), the faster the IP speech rate becomes ($p=0.0325$*). The predictor NO_OFL in the model is significant ($X^2(4.61)=1$, $p=0.032$*). None of the other predictor variables and their interactions returned significant results. Therefore, H0 for RQ11d has to be accepted, since no stay abroad effects were found.

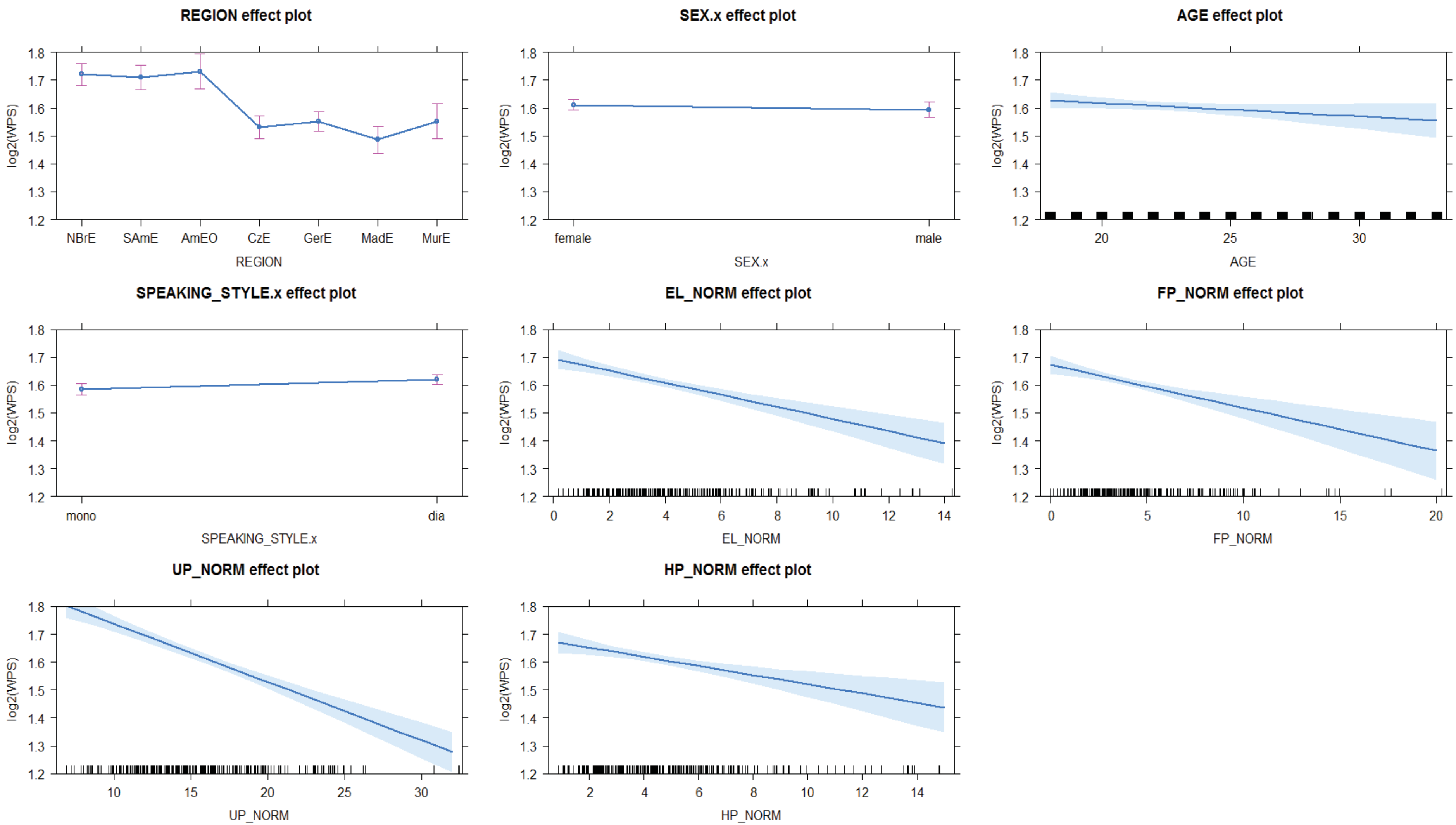


Figure 24. Effect plots for the mixed effects model results of IP speech rate in WPS

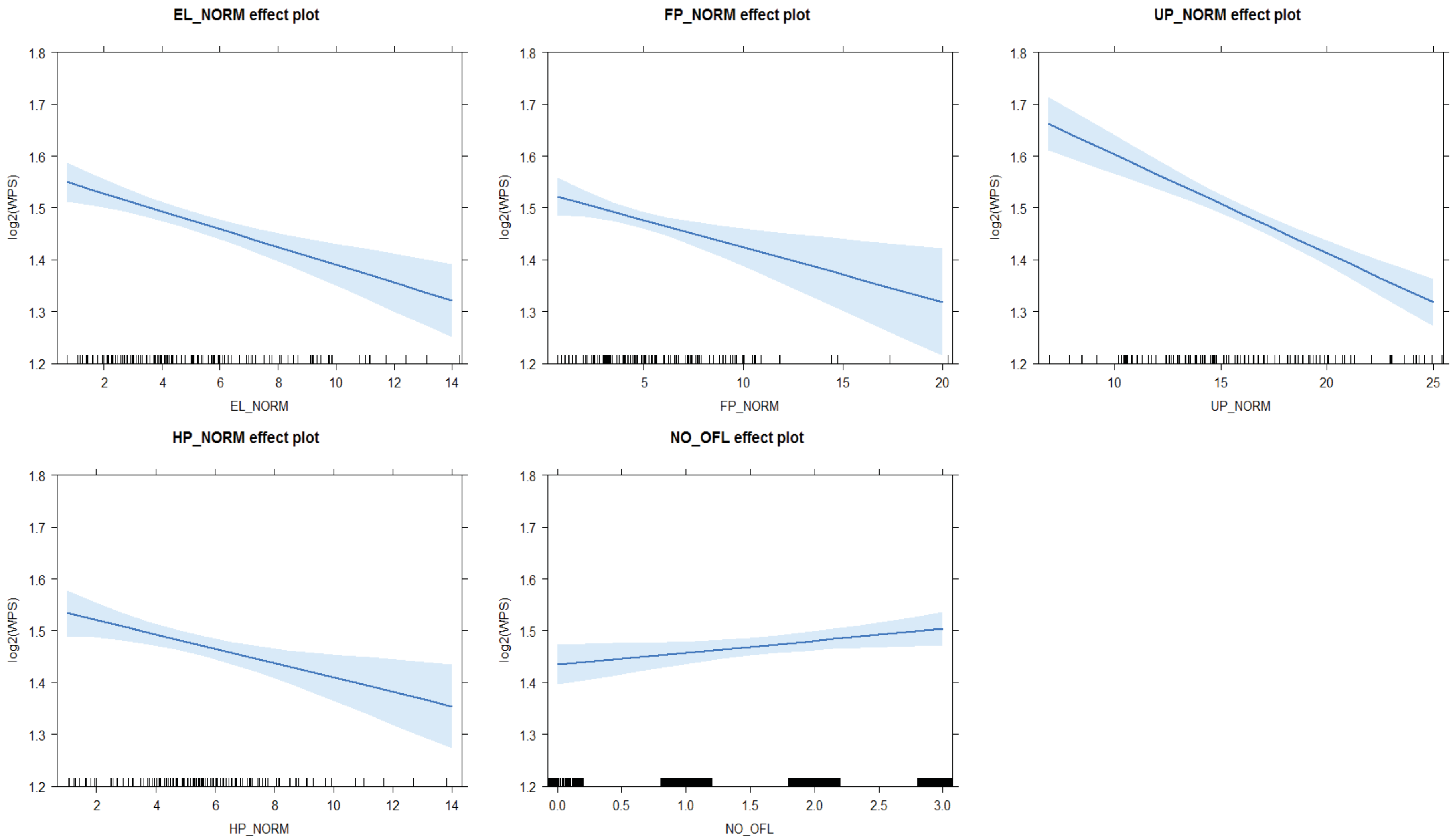


Figure 25. Effect plots for the mixed effects model results of WPS per IP for the complete learner data only (main effects only)

4.2.2 ip measures

A summary of the descriptive statistics of ip frequency, length (seconds, words), and speech rate per ip (in WPS) is given in Table 22.

Table 22. Descriptive statistics for ips, **absolute**/normalized in phw **(dispersion)**

Measures of central tendency	**NBrE**	**SAmE**	**AmEO**	**CzE**	**GerE**	**MadE**	**MurE**	Measures of dispersion in brackets
total number of ips ips phw	3,637 649	3,876 South: 2,646 471	3,876 Other: 1,230 220	3,200 571	3,764 671	4,164 Mad: 2,337 413	4,164 Mur: 1,827 363	
median ips/ips phw by speakers	82/**15** (10/1.7)	93/**17** (21/3.7)	76/**14** (12/1.9)	70/**13** (13/2.5)	81/**14** (13/2.4)	92/**16** (15/2.8)	92/**18** (18/2.4)	**IQR**
min-max ip/min-max ip_phw	60-94/ **10.3-17**	63-116/ **11.2-20**	57-103/ **10.2-18**	49-91/ **8.8-16**	53-124/ **9.4-22**	74-111/ **13.4-20**	43[75]-123/ **14.9-24**	
median length (s)/WPS	1.53/**4.17** (1.1/1.8)	1.47/**3.83** (1.1/1.8)	1.65/**4.07** (1.3/1.7)	2.43/**2.89** (2/1.2)	1.91/**3.12** (1.5/1.3)	1.73/**3.14** (1.3/1.4)	1.79/**2.78** (1.5/1.4)	**IQR**
min-max length (s)/WPS	0.24-9.35/**0.75-16.61**	0.24-8.14/ **0.78-34.33**	0.23-7/ **0.68-12.71**	0.25-13.3/**0.84-7.93**	0.27-13.25/**0.62-24.73**	0.3-8.82/ **0.3-9.71**	0.27-12.55/ **0.46-8.24**	
median words/WPS	6/**4.17** (5/1.8)	6/**3.83** (4/1.8)	6/**4.07** (5/1.7)	7/**2.89** (5/1.2)	6/**3.12** (5/1.3)	5/**3.14** (4/1.4)	5/**2.78** (4/1.4)	**IQR**
min-max words/ WPS	1-26/ **0.75-16.61**	1-38/ **0.78-34.33**	1-33/ **0.68-12.71**	1-34/ **0.84-7.93**	1-38/ **0.62-24.73**	1-24/ **0.3-9.71**	1-23/ **0.46-8.24**	
median of tones per ip	2 (1)	2(1)	2 (1)	3 (1)	3 (1)	2 (1)	2 (1)	**IQR**
min-max tones per ip	2-5	2-5	2-5	2-9	2-7	2-6	2-7	

[75] This low score is due to the speakers with 250 words samples (n=2).

In previous research one of the factors that was identified to influence intonational phrasing and pausing behavior was the use of glottalization (creaky voice), which was found to be frequent in Czech and German L2 speech (Bissiri and Volín 2010; Bissiri 2013). In order to explain the ips better, instances where speakers adopted CV have to be investigated as well. Because, if learners really use CV more frequently, which is associated with phrase boundaries, then this would mean that they may produce more ips, and/or the annotators possibly used this segmental feature as a cue to intonation phrase breaks. The overview in Table 23 and the following linear regression include also stretches that were excluded from the f0 range analysis in Section 4.3 due to unreliable f0 contours caused by the CVs. Minor CV instances were left in and are analyzed in the co-occurrence with tonal targets in the regional varieties of English. The mean frequency of CV instances in the different varieties of English are broken down in Table 23.

Table 23. Descriptive statistics for creaky voice instances in the selected samples by REGION

	REGION	**Mean**	**SD**	**Range**
1	NBrE	4.87	4.48	0-16
2	SAmE	22.69	17.83	1-73
3	AmEO	11.69	10.38	0-39
4	CzE	9.42	8.77	0-38
5	GerE	9.87	8.39	1-37
6	MadE	7.12	6.73	0-31
7	MurE	9.05	6.52	1-27

These CV instances can be of varying lengths and span various tonal targets. The CV instances mostly co-occur with low edge tones (L-, and L%), as illustrated in Figure 64 (see Appendix). 75% of the NBrE CV instances appear with low edge tones, 78% in SAmE, and 79% in the mixed AmEO group. The learner productions get quite close to these values with 71% for GerE, MadE 71%, and MurE 67.5%. For the Czech learners, however, only 50% of the CVs co-occur with low edge tones. Most NSs do not use CV with high edge tones. The CzE and MurE speakers, however, both produce 11% of their CVs on high edge tones. CVs on low pitch accents are also quite infrequent but SAmE and MadE speakers produce 14% of their low pitch accents with CVs. AmEO and the GerE speakers produce 8% of low pitch accents with a CV. High pitch accents are very rarely produced with CVs, with a median of 0% for all regional varieties, but the MurE speakers produce it 8.8% of the time. Since the mean range of NE overall CV production is between 4-22 instances, the learners can be said to be very well within those limits. However, when it comes to the type of boundary tones, some learners also use CV with high boundary tones, while NSs almost never do.

Since only one data point per speaker and their CV frequency was given, a linear regression was fit. The linear regression (log2(CV_TOTAL) ~ 1 + REGION + SEX + log2(INTERRUPTIONS)) ($R^{2=}$0.21, F=8.26 on 8 and 216 DF, p<0.001***) determined that NBrE speakers use CVs the least and SAmE speakers the most (see Table 24). Differences between NBrE and SAmE are larger than differences between native and non-native speakers. Learners, however, all produce significantly more CV instances than the NBrE speakers and less than SAmE speakers and they are overall quite similar to each other. In general, male speakers produce fewer CV instances and the more interruptions by the interviewer there are, i.e. the more dialogic the speech is, the more CV instances occur. REGION is the most important factor to explain the distribution of CV (p<0.001***), followed by SEX (p=0.00119**), and INTERRUPTIONS (p=0.04369*). The explanatory power of the model, however, is not satisfactory, only explaining about 21% of the variation in the data (adjusted $R^{2=}0.21$).

Table 24. Results for linear regression with creaky voice frequency. Final model: log2(CV_TOTAL) ~ 1 + REGION + SEX + log2(INTERRUPTIONS)

Predictors	**Estimate**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	1.8183	0.3198	5.69	< 0.001	***
REGION_SAmE	2.2284	0.326	6.83	< 0.001	***
REGION_AmEO	1.3859	0.373	3.72	< 0.001	***
REGION_CzE	0.8013	0.285	2.81	0.00538	**
REGION_GerE	1.0466	0.2661	3.93	< 0.001	***
REGION_MadE	0.471	0.3134	1.5	0.13439	
REGION_MurE	0.9278	0.3311	2.8	0.00554	**
SEX_male	-0.6081	0.1851	-3.29	0.00119	**
INTERRUPTIONS_log	0.1580	0.0779	2.03	0.04369	*

The results of the linear regression are summarized in the effect plots in Figure 26.

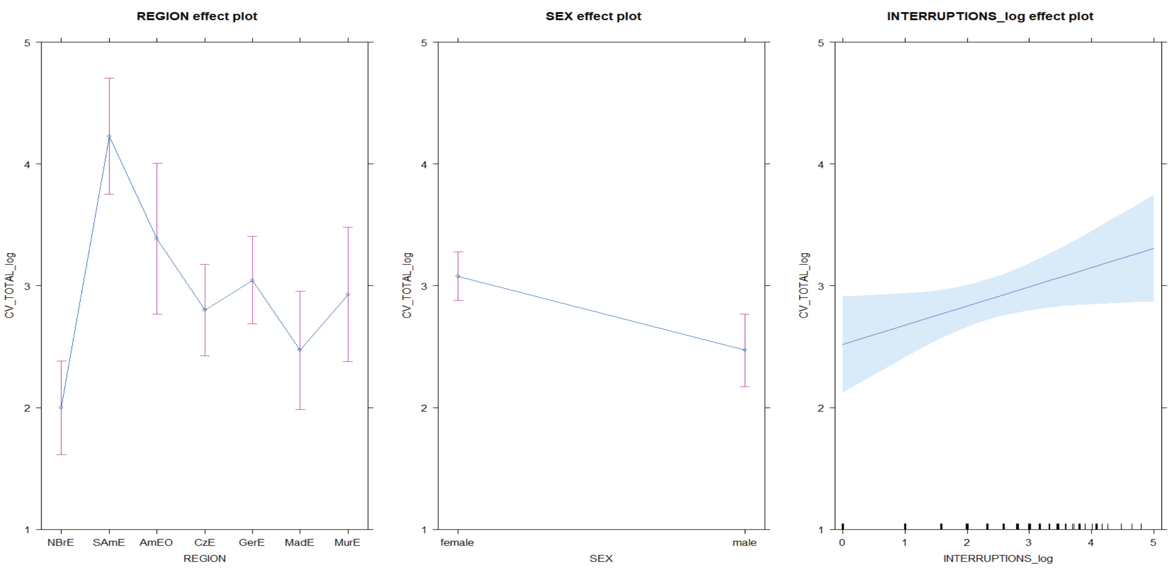


Figure 26. Effect plots for linear regression results of creaky voice instances

These results confirm previous research that also found that GerE and CzE speakers use CV more often than the BrE native controls (Bissiri and Volín 2010; Bissiri 2013). This result could be seen as evidence for L1 influence (cf. Bissiri et al. 2011). However, if the AmE data is considered, this is rather unlikely to be due to an L1 influence, especially since all the learners have extremely similar distributions. The learners in this study use intermediate values between NBrE and AmE speakers, and approach none of the targets (except MadE speakers). As predicted by a previous study by Bissiri et al. (2011), Spanish NSs use CVs quite infrequently. Only the MadE speakers approach NBrE values and the difference between the two groups is not significant. Barry (2008: 125) also found a considerable amount of CV instances in SoCal speakers, compared to BrE speakers from London, which was used either with particularly high or particularly low pitch. A separate linear regression with the learner data only, revealed no significant differences between the learner varieties, and no significant change with a stay abroad, years of English at university and school, and age. Thus, the results by Bissiri (2013: 236) with CV phenomena becoming less frequent and more native-like with longer stays abroad could not be verified in the present study. In sum, even though significant difference between NBrE and the learner varieties were found, the learner varieties seem to be within the limits, taking intermediate values between NBrE and SAmE speakers. Since even NSs diverge so much from each other, CV instances are not of a concern for the learners in this study. The learners, however, do produce CV instances less on low edge tonal categories and also on high tones, compared to the NSs. Overall, the use of CV does not explain the production of more ips by the learners, which will be explained in the next section.

4.2.2.1 ip frequency

In total there are 18,641 ips in all five corpora/seven regional varieties of English. 10,774 ips were uttered within the dialogues and 7,867 within the monologues. 13,031 ips were uttered by female speakers and 5,610 by male speakers. Visual inspection with the help of ggplots2 determined that there does not seem to be a large difference in ip frequency between the seven different varieties of English, when the values are normalized and logged. In order to answer RQ3, various linear regressions were built that included all of the fluency variables and segmental variables (creaky voice) that were thought to influence ip frequency. A linear regression was selected for this research question because there was only one data point per speaker, i.e. the specific number of ips per speaker. RQ3: *Do learners of English produce more prosodic breaks (ips) than NSs of English? Does this differ depending on the speaking style, region, age, sex, and fluency (measures in WPS per ip, number of silent and filled pauses, number of hesitations, number of elongations)?*

H0	the distribution of the DV ip frequency (IM_PHW) of the NNSs equals that of the NSs, and does not differ depending on the levels of the IVs SPEAKING_STYLE, REGION, SEX, AGE, UP, FP, HP, EL, and WPS_IM: $R^2=0$.
H1	the distribution of the DV ip frequency (IM_PHW) of the NNSs is larger than that of the NSs, and differs depending on the levels of the IVs SPEAKING_STYLE, REGION, SEX, AGE, UP, FP, HP, EL, and WPS_IM: $R^2>0$.

The final model ($R^2=0.66$, F=32.2 on 14 and 210 DF, $p<0.001$) only included the independent variable UP_NORM (filled pauses) and the interaction between REGION * SEX. Table 25 shows the results of the linear regression:

Table 25. Results for linear regression with ips phw. Final model: log2(IM_PHW) ~ 1 + REGION * SEX + log2(UP_NORM)

Predictors	Estimate	Std. Error	t-value	p-value	
(Intercept)	2.56185	0.10418	24.59	<0.001	***
REGION_SAmE	0.13802	0.04597	3	0.003	**
REGION_AmEO	-0.0154	0.0663	-0.23	0.8165	
REGION_CzE	-0.29893	0.03885	-7.69	<0.001	***
REGION_GerE	-0.06308	0.04075	-1.55	0.1231	
REGION_MadE	-0.00238	0.04716	-0.05	0.9598	
REGION_MurE	0.04434	0.05683	0.78	0.4362	
SEX_male	-0.03972	0.04448	-0.89	0.3729	
log2(UP_NORM)	0.35527	0.02842	12.5	<0.001	***
REGION_SAmE:SEX_male	-0.19011	0.07285	-2.61	0.0097	**
REGION_AmEO:SEX_male	-0.073	0.08743	-0.83	0.4047	
REGION_CzE:SEX_male	0.13593	0.07886	1.72	0.0862	.
REGION_GerE:SEX_male	-0.03388	0.06534	-0.52	0.6046	
REGION_MadE:SEX_male	0.08365	0.08537	0.98	0.3283	
REGION_MurE:SEX_male	-0.09131	0.08429	-1.08	0.2799	

Together the variables explain 66% of the variance in the DV (R^2=0.66). The final linear regression (log2(IM_PHW) ~ 1 + REGION * SEX + log2(UP_NORM)) revealed that female SAmE speakers are predicted to produce significantly more ips (p=0.003**), while female CzE speakers (p<0.001***) and SAmE males (p=0.0097**) are predicted to produce significantly less. Also, the more silent pauses are produced, the more prosodic breaks are produced (p<0.001***). Visual inspection of residual plots did not show any deviations from normality. Multicollinearity was tested and was determined to be negative. Just as the IP frequency linear regression revealed, the interaction between REGION and SEX is significant (p=0.0062**), as well as the frequency of UPs (p<0.001***). However, in the ip frequency analysis SPEAKING_STYLE does not have an effect. Generally, the R^2 value is higher for this model (compared to the IP_PHW model) by 14%, because ip frequency is less variable than IP frequency. The results of the linear regression are depicted in the effects plot in Figure 27.

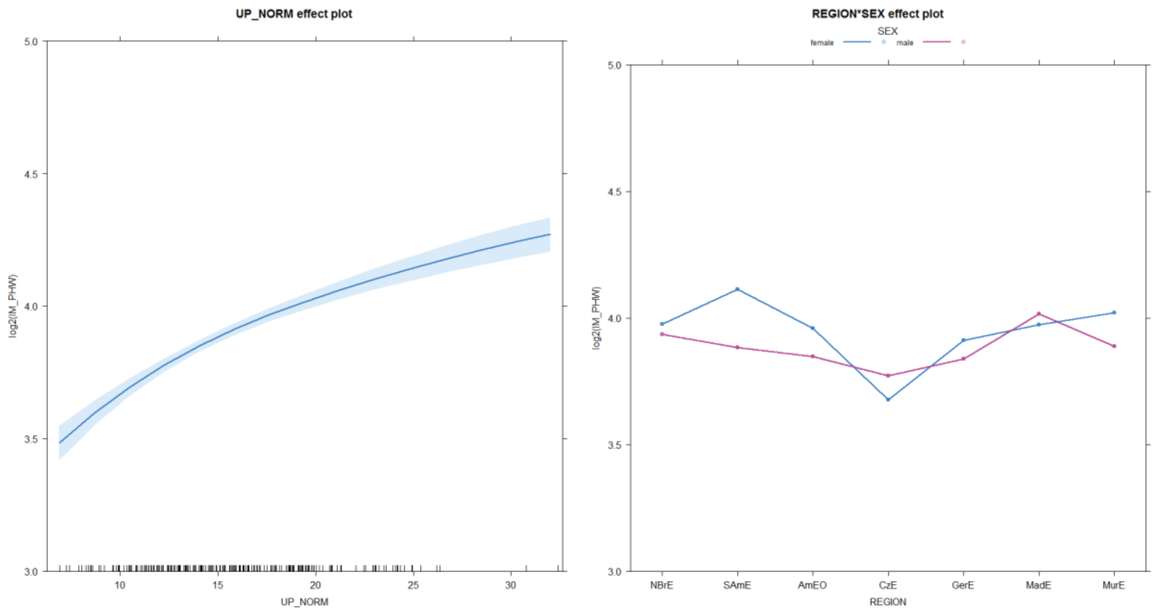


Figure 27. Effect plots for linear regression results of ips phw

However, one could argue that silent pauses are the most common indicators for prosodic breaks at the ip level and, therefore, this could only be logical for these two variables to co-occur. Thus, if silent pauses are left out of the linear regression, the R^2 value drops to 0.41 (R^2=0.41, F=13.1 on 13 and 211 DF, p<0.001***), and both Spanish groups are predicted to produce significantly more ips than the NBrE intercept (see Figure 28). The results for the CzE and SAmE speakers remain unchanged.

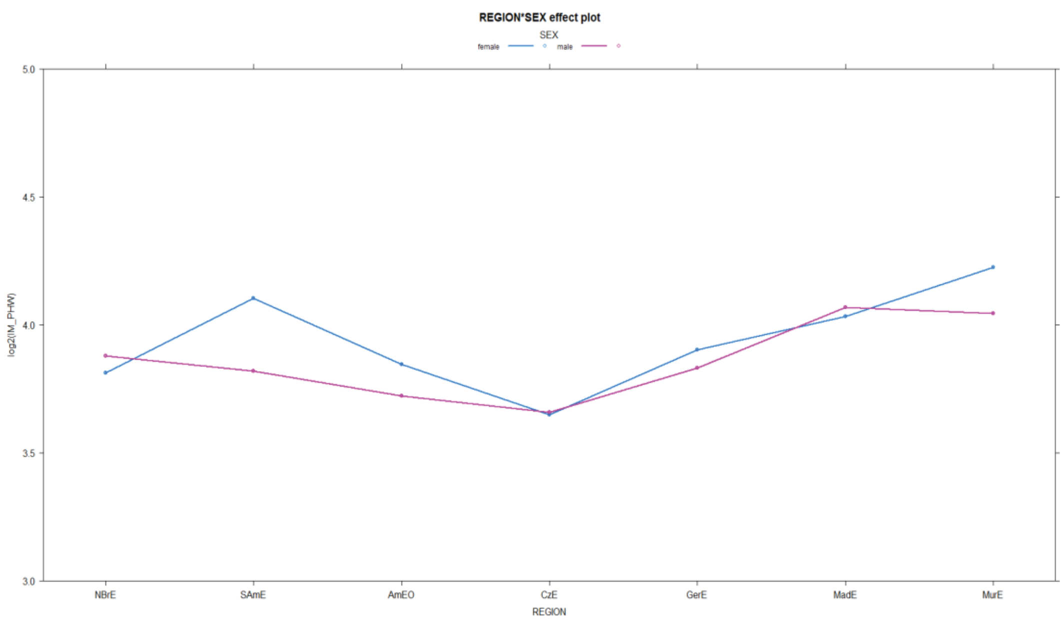


Figure 28. Effect plots for linear regression results of ips phw without UP_NORM

Thus, if silent pauses are not taken into account in the linear regression, H1 of RQ3 is true only for Spanish speakers. However, if silent pauses are taken into account, H1 can be rejected for all NNSs, since no differences were found for any of the learners, except the CzE female speakers, which, however, go in a different direction than expected, i.e. they produce significantly less ips than the NBrE intercept.

Another model (log2(IM_PHW) ~ 1 + REGION + log2(UP_NORM) + filled_pauses_log + REGION*YOEAU, R^{2}=0.71, F=40.1 on 7 and 107 DF, p<0.001***) was fit (results are shown in the effect plots in Figure 29), containing only the complete data of the CzE, GerE, and MadE speakers' proficiency and stay abroad information, i.e. STAB, YOEAS, YOEAU, the NO_OFL, and AGE, as well as the variables from the previous analysis to determine ip frequency.

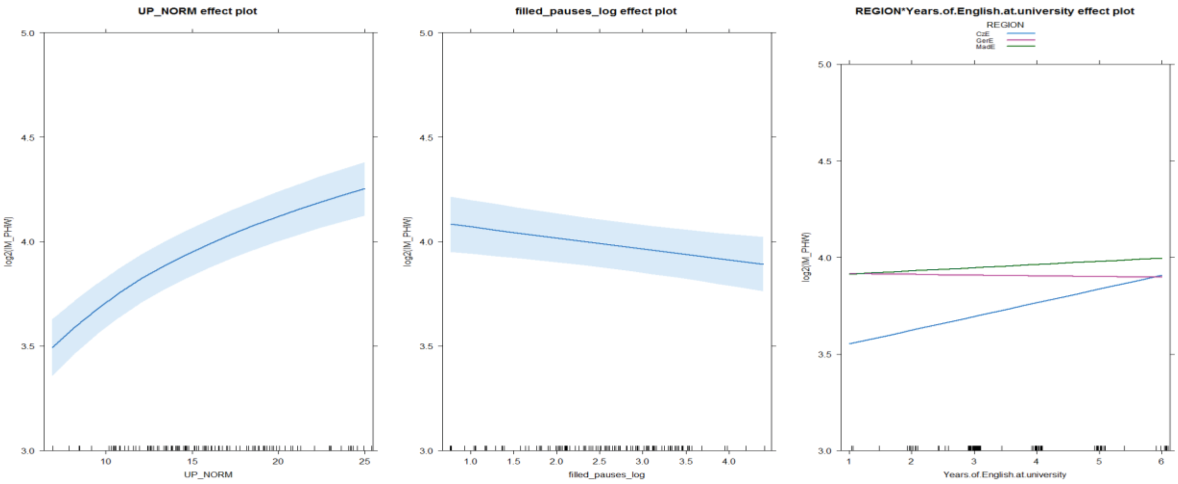


Figure 29. Effect plots for linear regression results of ips phw for the complete learner data only

Only the predictors of FP (p=0.0042***) and UP (p<0.001***) and one interaction REGION*YOEAU (p=0.0167*) significantly contributed to the predictive power of the model. The linear regression revealed that the GerE (p<0.001***) and the MadE speakers (p<0.001***) differ significantly from the CzE speakers, in that they produce significantly more ips phw. It was also revealed that the more silent pauses are used, the more ips are predicted to be produced (p<0.001***) (Plot 1 in Figure 29), and the more filled pauses are used, the fewer breaks are made (p=0.00422**) (Plot 2). These results seem to point to two opposing strategies in the learner data, where silent pauses seem to function as disjuncture and filled pauses as utterance-binding features. Nevertheless, filled pauses are frequently found to occur at syntactic junctions between clauses, phrases, and semantically heavy words (Kjellmer 2003), which together with the prosodic strength at these junctions may be taken as either IU-binding or -separating features. While years of English at university does not have a significant effect on the GerE and MadE speakers, the CzE speakers seem to increase their ip frequency to the level of the GerE speakers (who were not significantly different in their production of ips phw from the NBrE speakers) with an increasing number of years of English at university (Plot 3). Thus, the longer the CzE speakers were at university, the higher their ip frequency gets (p=0.00196**). For RQ11c (*Do ip breaks become less frequent after a stay abroad?*) it can be said that the STAB does not improve model fit, and neither does the interaction between REGION:STAB.

4.2.2.2 ip speech rate in WPS

NSs produce between 57-116 (10-20 in ip phw) ips per 560 words. The NNSs produce between 43-124 (8-24 in ip phw) ips per 560 words. Figure 65 (see Appendix) shows that

NNSs produce a slower speech rate per ips (measured in WPS) compared to the three NS groups. The following mixed effects model was intended to answer RQ4: *Do learners of English produce slower speech rates in ips compared to NSs? Does this differ depending on speaker sex, age, speaking style, and fluency?*

H0 the ip speech rate (WPS) of the NNSs is equal to that of the NSs, and does not differ depending on the levels of the IVs SEX, AGE, SPEAKING_STYLE, INTERRUPTIONS, UP, and FP: $R^2 = 0$.

H1 the ip speech rate (WPS) of the NNSs is smaller than that of the NSs, and differs depending on the levels of the IVs SEX, AGE, SPEAKING_STYLE, INTERRUPTIONS, UP, and FP: $R^2 > 0$.

The same mixed effects model (conditional R^2=0.23) as for the IPs in Section 4.2.1.2 was used for ip speech rate. The only difference between the two models is that for the ips it is possible to add the independent variable of the pragmatic functions (PRAGM) of the ips. A by pragmatic function analysis for the IPs was not possible because each IP can contain ips with different pragmatic functions. However, each ip was assigned only one primary pragmatic function. Results of the mixed effects model are reported in Table 26:

Table 26. Results for mixed effects model with WPS per ip. Final model: log2(WPS_IM) ~ REGION + SEX + AGE + SPEAKING_STYLE + EL_NORM + FP_NORM + UP_NORM + HP_NORM + PRAGM + (1 | SPKR). Speech rate measures are in words per seconds (WPS)

Fixed effects:	**Estimate**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	2.48602	0.07313	34	< 0.001	***
REGION_SAmE	0.00395	0.02824	0.14	0.88897	
REGION_AmEO	-0.00411	0.03488	-0.12	0.90631	
REGION_CzE	-0.22378	0.03133	-7.14	< 0.001	***
REGION_GerE	-0.18907	0.028	-6.75	< 0.001	***
REGION_MadE	-0.21417	0.03259	-6.57	< 0.001	***
REGION_MurE	-0.1612	0.04146	-3.89	< 0.001	***
SEX_male	0.00309	0.01801	0.17	0.86405	
AGE	-0.0074	0.00286	-2.59	0.01025	*
SPEAKING_STYLE_dia	0.01168	0.01037	1.13	0.26007	
EL_NORM	-0.02209	0.00374	-5.91	< 0.001	***
FP_NORM	-0.01397	0.00335	-4.17	< 0.001	***
UP_NORM	-0.01745	0.00221	-7.91	< 0.001	***
HP_NORM	-0.01122	0.00441	-2.54	0.01168	*
PRAGM_checking	0.02018	0.04064	0.5	0.61956	
PRAGM_emphasis	-0.0842	0.07251	-1.16	0.24559	
PRAGM_finality	0.07907	0.0234	3.38	< 0.001	***
PRAGM_handing over turn	0.05889	0.01474	4	< 0.001	***
PRAGM_insecurity	-0.15991	0.04377	-3.65	< 0.001	***
PRAGM_listing	-0.18473	0.01371	-13.48	< 0.001	***

Number of obs.=18,561, Random intercepts: Speaker SD (0.0991), Residual SD (0.4745); Log likelihood =-12658, Conditional R^2=0.24.

The results are visualized in the effect plots in Figure 30. The NBrE female intercept in monologic speech in continuation utterances (pragmatic function) produce a speech rate of 2.54 WPS (on a logged scale) per ip. All other regional varieties of English produce a slower speech rate (except SAmE) but the difference is only highly significant for all NNSs (Plot 1 in Figure 30). The ip speech rate is slowest in descending order in CzE, MadE, then GerE, and finally MurE, the other factors being accounted for. The AmEO group also produce a slower speech rate per ip, but the difference is not significant. Overall the predictor REGION reached a highly significant effect ($X^2(6)=97.9$, $p<0.001$***). Male speakers tend to produce a faster speech rate but the difference is not significant (Plot 2). The older speakers are, the slower their ip speech rate becomes ($p=0.0026$**) (Plot 3). In dialogues, ip speech rate is faster, but the result is not significant (Plot 4). The higher the number of fluency features such as elongations, silent and filled pauses, and hesitations all contribute to a slower speech rate (Plots 5-8). The pragmatic functions influence ip speech rate as well, with ips signaling finality (FIN) and handing over of turns (HOT) tending to involve a faster speech rate and those signaling insecurity (INSEC) and listing (LIST) tending to be produced with a slower speech rate (all $p<0.001$***) (Plot 9). Overall, however, all variables together explain only 23% of the variance in the DV, with a large portion of variance left unaccounted for. As with the speech rate for IPs, speech rate within ips is significantly generally slower in non-native speech and therefore I can decide in favor of H1 of RQ3.

Another mixed effects model was fit (conditional $R^2=0.092$), containing only the complete data of the CzE, GerE, and MadE speakers' proficiency and stay abroad information, i.e. STAB, years of English at school and university and the age of the learners, as well as the variables from the previous analysis to determine ip speech rate, thus answering RQ11d: *Do learners produce a faster speech rate (WPS) per ip after a stay abroad?* The mixed effects model did not yield any significant results for the proficiency or stay abroad variables. Also, no significant differences between the learner groups could be observed. Only fluency variables (EL, FP, and UP) showed significant effects in the same direction as the previous model.

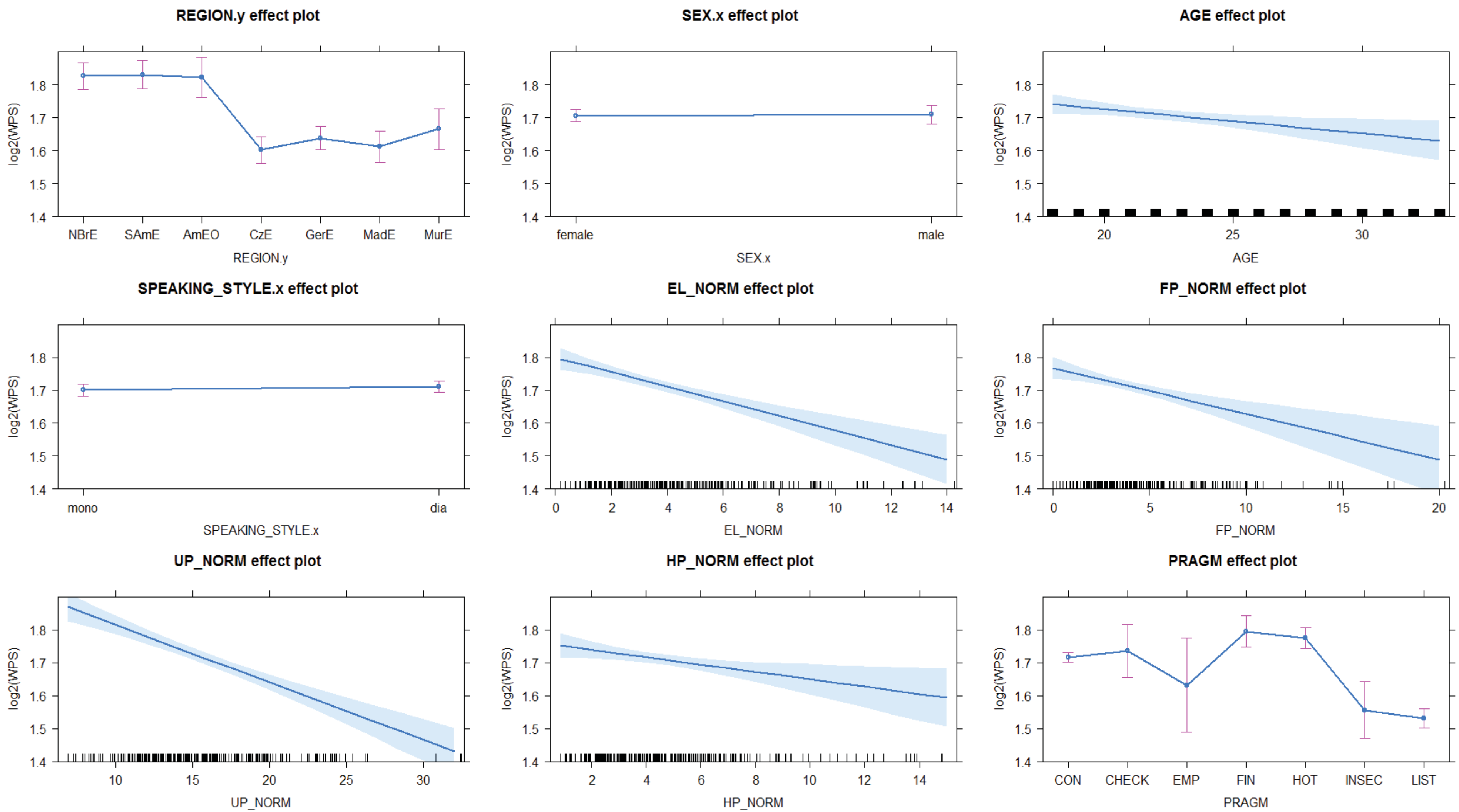


Figure 30. Effect plots for mixed effects model results for ip speech rate in WPS

4.3 F0 range in native and non-native speech

In order to be able to compare results of previous studies with the present study, I adopted Patterson's (2000) approach and included f0 range measures of level and span, long-term distribution (LTD) measures per speaker, and linguistic measures of the tone labels. In order to test the differences in f0 height (span and level), f0 maxima and minima across the different L1 groups, linear mixed-effects regression models were employed. The TUNE_PATTERN_ confl variable in this section includes four different types of tunes: HH, LL, HL, and LH. These were conflated because in the previous Section (4.1.6) on tune choice, a more detailed by-tune analysis was undertaken, where I distinguished between simple and complex tunes and narrow and wide f0 range tunes (HH vs. HF and LL vs. LF). For the f0 range analysis these tunes were conflated, i.e. all tunes include also the complex tunes where a tune might be repeated several times but no ip break is present. They were grouped according to their general direction or change of a ToBI label. HL, for instance, has a high starting point and a low end point, and it does not matter how many times the pattern is repeated in between. While a distinction of narrow and wide f0 tune, e.g. HH vs. HF would lead to logical dependencies in the f0 range analysis, the tunes were conflated, because if a tune retains the same tonal categories throughout that does not mean that they are necessarily flat. HH tunes can have just as a wide f0 levels and spans as HL tunes.

4.3.1 F0 range by tone labels

The present section describes the f0 realization of ToBI categories. Table 27 gives a summary of f0 median by TONE LABEL, REGION, and SEX.

Table 27. Median f0 by TONE LABEL, REGION, and SEX

FEMALE					MALE				
REGION	**LABEL**	**F0**	**LABEL**	**F0**	**REGION**	**LABEL**	**F0**	**LABEL**	**F0**
NBrE	%H	258	L*	177	**NBrE**	%H	172.5	L*	105
SAmE	%H	289	L*	176	**SAmE**	%H	125	L*	114
AmEO	%H	309	L*	181.5	**AmEO**	%H	176	L*	92
CzE	%H	237	L*	181	**CzE**	%H	141.5	L*	125
GerE	%H	257	L*	185	**GerE**	%H	162	L*	106
MadE	%H	262.5	L*	183	**MadE**	%H	144	L*	103
MurE	%H	264.5	L*	179	**MurE**	%H	173	L*	121.5
NBrE	H*	218	L-	175	**NBrE**	H*	127	L-	103
SAmE	H*	237	L-	171	**SAmE**	H*	139	L-	110
AmEO	H*	240	L-	177	**AmEO**	H*	127	L-	94
CzE	H*	214	L-	173	**CzE**	H*	136	L-	123
GerE	H*	226	L-	181	**GerE**	H*	126	L-	105
MadE	H*	221	L-	181	**MadE**	H*	115	L-	101
MurE	H*	206	L-	171	**MurE**	H*	139	L-	119
NBrE	H-	200	L%	167	**NBrE**	H-	116	L%	102
SAmE	H-	224	L%	167	**SAmE**	H-	128	L%	111
AmEO	H-	225	L%	174	**AmEO**	H-	119	L%	92

FEMALE					MALE				
REGION	**LABEL**	**F0**	**LABEL**	**F0**	**REGION**	**LABEL**	**F0**	**LABEL**	**F0**
CzE	H-	207	L%	169	**CzE**	H-	128	L%	119
GerE	H-	220	L%	175	**GerE**	H-	121	L%	104
MadE	H-	220.5	L%	180	**MadE**	H-	115	L%	96
MurE	H-	214	L%	170	**MurE**	H-	131	L%	118
NBrE	H%	195.5			**NBrE**	H%	117		
SAmE	H%	215			**SAmE**	H%	120		
AmEO	H%	219			**AmEO**	H%	117		
CzE	H%	220			**CzE**	H%	139		
GerE	H%	234.5			**GerE**	H%	132		
MadE	H%	221			**MadE**	H%	121		
MurE	H%	213			**MurE**	H%	135		

The results are additionally visualized in the following ggplots in Figure 66 (see Appendix), which show that female AmE speakers have the highest starting points (%H) and the lowest end-points (L%). The female NBrE speakers are the ones with the flattest f0 over tone labels, especially in H% boundary tones. The GerE speakers stick out with the highest final high boundary tone (H%) f0 values. However, all female speakers produce similar f0 values for low tones (L*, L-, and L%). The ggplots in Figure 67 (see Appendix) show that for both female and male NSs start their utterances with higher f0 values and end them lower, while the learners all produce higher f0 values for high boundary tones.

Another aspect that has to be accounted for when analyzing f0 is whether the speakers use quoted or non-quoted speech, because previous research has shown that quoted speech is generally higher in f0 than non-quoted speech and it is frequently used in narrative speaking styles (e.g. Jansen et al. 2001; Tyler 2014). The distribution of quoted speech samples by all speakers is demonstrated in Table 28.

Table 28. Distribution of quoted speech samples by speaker, REGION, and SEX

SPKR	SEX	Freq	SPKR	SEX	Freq	SPKR	SEX	Freq	SPKR	SEX	Freq
AE003	f	3	**CZ004**	f	1	**SP007**	f	2	**AE021**	m	2
AE004	f	9	**CZ005**	f	1	**SP008**	f	2	**AE029**	m	12
AE006	f	4	**CZ009**	f	1	**SP011**	f	4	**AE030**	m	1
AE007	f	4	**CZ011**	f	1	**SP017**	f	4	**AE033**	m	3
AE010	f	7	**CZ014**	f	4	**SP019**	f	8	**AE042**	m	1
AE011	f	4	**CZ026**	f	5	**SP022**	f	3	**BE007**	m	2
AE014	f	3	**CZ029**	f	1	**SP025**	f	5	**BE014**	m	2
AE016	f	7	**CZ030**	f	1	**SP027**	f	1	**BE017**	m	5
AE017	f	5	**CZ032**	f	3	**SP030**	f	5	**BE019**	m	1
AE049	f	1	**GE001**	f	3	**SP031**	f	5	**BE026**	m	4
BE008	f	2	**GE007**	f	3	**SP035**	f	1	**BE027**	m	2
BE009	f	8	**GE009**	f	1	**SP037**	f	2	**BE038**	m	3
BE010	f	2	**GE016**	f	1	**SP040**	f	1	**BE043**	m	1
BE015	f	1	**GE022**	f	1	**SP047**	f	1	**BE049**	m	4

SPKR	SEX	Freq	SPKR	SEX	Freq	SPKR	SEX	Freq	SPKR	SEX	Freq
BE016	f	1	**GE025**	f	1				**CZ012**	m	2
BE022	f	6	**GE027**	f	6				**CZ049**	m	1
BE024	f	1	**GE028**	f	1				**GE004**	m	3
BE033	f	3	**GE032**	f	4				**GE010**	m	1
BE046	f	6	**GE042**	f	3				**SP001**	m	4
BE048	f	2	**GE049**	f	1				**SP005**	m	6
CZ001	f	2	**SP004**	f	2				**SP028**	m	2
CZ003	f	3	**SP006**	f	1				**SP049**	m	2

80 speakers out of a total of 225 use quoted speech in the respective samples. In the selected corpus sample there are 238 instances of quoted speech. This is labeled as "imit" in my corpus and includes not only imitations of speakers but also quoted speech. The quoted phrases (by ip) are listed by L1 below. For most regional varieties, the quoted phrases comprise 1-2% of the total phrases per group.

AmE	66	from 15 different speakers (5 male and 10 female)
SpE	61	from 20 different speakers (4 male and 16 female)
		(Murcia females: 16, males: 4, Madrid females: 31, males: 10)
NBrE	56	from 19 different speakers (9 male and 10 female)
GerE	29	from 13 different speakers (2 male and 11 female)
CzE	26	from 13 different speakers (2 male and 11 female)

As described in Section 3.3 (Selected recordings and speaker profiles), the non-native corpora are unbalanced as to speaker sex. There are only very few male speakers in the CzE, GerE, and SpE corpora. Therefore, normalized values are given in Table 29. The trend is similar in all corpora; quoted speech is more common in female speech (174 instances of quoted speech in the entire corpus stem from female speech, compared to 64 from male speech). Similar results have been found by Jansen et al. (2001). For the MadE males, quoted phrases are just as frequent as those of AmE females. The CzE and GerE speakers and the speakers from the NWSP corpus use quoted speech quite infrequently. Overall, the AmE, NBrE, and SpE speakers come very close to the overall occurrence of quoted phrases in the data set. This could be explained by more informal speaking styles, or the perception of a more informal situation (see discussion on formality in Section 3.8).

Table 29. Quoted speech phrases freq. by REGION and SPEAKING STYLE. Normalized values=quoted speech freq. by REGION/total ip frequency by REGION * 100)

No.	REGION	SPEAKING STYLE	quoted_freq	total ip freq	norm_freq
1	NBrE	dia	34	2,725	1.25
2	SAmE	dia	10	1,099	0.91
3	AmEO	dia	1	653	0.15
4	CzE	dia	10	1,691	0.59
5	GerE	dia	16	2,108	0.76
6	MadE	dia	32	1,353	2.37
7	MurE	dia	14	1,145	1.22

No.	REGION	SPEAKING STYLE	quoted_freq	total ip freq	norm_freq
8	NBrE	mono	22	912	2.41
9	SAmE	mono	40	1,547	2.59
10	AmEO	mono	15	577	2.60
11	CzE	mono	16	1,509	1.06
12	GerE	mono	13	1,656	0.79
13	MadE	mono	9	984	0.91
14	MurE	mono	6	682	0.88

In the NS corpora quoted phrases are more frequently used in monologues. However, in the dialogues from MadE quoted phrases are twice as frequent in dialogues than in monologues. Interestingly, quoted phrases from the SAmE corpus can be predominantly found in monologues. The SAmE files are highly narrative and informal in nature and this could explain the more frequent occurrence of quoted speech in general. However, in normalized values, the NSs all have extremely similar occurrences of quoted speech in monologues.

Now that the frequencies of quoted speech samples have been determined, I will turn to the analysis of f0 range within these quoted speech samples. Speakers who have an unusually high f0 for quoted speech, which is out of their normal non-quoted f0 level are: BE033, BE038, BE046, BE048, AE010, AE021, AE30, MurE: SP30, SP31, SP037, SP047, and MadE: SP006, SP007, SP011, SP019, SP025. The CzE and GerE corpora include only very few speakers who have a higher f0 level for quoted speech than for non-quoted speech: GE007, CZ001, CZ005, CZ012 and the difference between quoted and non-quoted speech is not as dramatic, as for those SpE, NBrE, and AmE speakers.

As one can see from Figure 68 (see Appendix), non-quoted speech samples in both sexes show more similarities across regional varieties than the quoted utterances, because they almost form one line on the plot. High initial boundary tones (%H) do not occur in quoted speech samples because the quoted samples never occurred immediately at the beginning of IPs. The quoted samples from the female speakers show greater f0 variability for the high-register ips with H* H-H% patterns. Especially the female SpE and GerE speakers seem to have an exaggerated f0 compared to their non-quoted speech samples. The CzE speakers show no larger differences from quoted to non-quoted speech in both sexes. The NSs do not seem to change their f0 much in quoted speech.

The descriptive statistics so far do not give a complete picture of the f0 variability in the data. Therefore, a mixed effects model was employed in order to answer RQ5: *Do learners of English produce native-like f0 levels? Does this differ depending on the speaking style, tune pattern, sex, age, number of topic initiations, position in the IP, ip length, pragmatic function, number of interruptions, and hesitations?*

H0 the f0 level of the NNSs is higher or lower than that of the NSs, and differs depending on the levels of the IVs: REGION, TUNE_PATTERN, SEX, POSITION, SPEAKING_STYLE, PRAGM, INTERVIEWER_INFLUENCE, UPTALK; QUOTED, TOPIC_INITIATIONS, and IM_LENGTH: $R^2 = 0$.

H1 the f0 level of the NNSs is equal to that of the NSs, and does not differ depending on the levels of the IVs: REGION, TUNE_PATTERN, SEX, POSITION, SPEAKING_STYLE, PRAGM, INTERVIEWER_INFLUENCE, UPTALK; QUOTED, TOPIC_INITIATIONS, and IM_LENGTH: $R^2 > 0$.

When modeling the learners' use of f0 by tone, certain considerations have to be made. The f0 values were taken from each individual tone (n=47,442) from the corpus samples. A detailed ToBI label analysis was not possible due to data sparcity issues for some REGION, SEX, and TONE_LABEL combinations. Therefore, the conflated categories of High (H) and Low (L) were considered instead. All f0 values of zero were excluded and, as the residuals exhibited a deviation from normality toward the extremes, but a good fit toward the center, data points with residuals deviating by more than two standard deviations were excluded from the analysis. The following statistical analysis is, therefore, based on 44,153 observations. The interaction between REGION and TONE_TYPE (high vs. low tones), SEX (male vs. female), and POSITION (accent vs. edge tone) are expected to be highly correlated with certain f0 values. Uptalk and quoted speech samples are expected to have significantly higher f0 values. Also, if a lot of topic shifts take place (TOPIC_INITIATIONS), then higher overall f0 values can be expected, because speakers usually start a new topic with a higher f0 level. In total, 157 speakers out of 225 speak about more than one topic within the selected 560-word speech samples. Also, the overall ip length needs to be considered, because previous research suggests that there may be larger differences in longer ips, compared to shorter ones (cf. Ramírez Verdugo 2002: 130; Zimmerer at al. 2015). The results of the linear mixed model (conditional $R^{2=}$0.83) are shown in Table 30.

Table 30. Results for mixed effects model with absolute f0 level by tone label. Final model: F0 ~ REGION * HL + SEX + POSITION + SPEAKING_STYLE + PRAGM + INTERVIEWER_INFLUENCE + UPTALK + QUOTED + TOPIC_INITIATIONS + IM_LENGTH + (1 | SPKR)

Fixed effects	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	225.2706	8.0495	27.986	< 0.001	***
REGION_SAmE	13.436	4.8995	2.742	0.006585	**
REGION_AmEO	8.033	5.8754	1.367	0.172895	
REGION_CzE	11.5494	6.0412	1.912	0.057171	.
REGION_GerE	13.4785	4.6334	2.909	0.003986	**
REGION_MadE	10.9375	5.1721	2.115	0.035545	*
REGION_MurE	2.477	5.4928	0.451	0.652457	
HL_LOW	-28.2085	0.4997	-56.456	< 0.001	***
SEX_male	-79.3416	3.0399	-26.1	< 0.001	***
POSITION_EDGE	-4.9399	0.2212	-22.336	< 0.001	***
SPEAKING_STYLE_dia	-0.3753	0.3522	-1.066	0.286581	
PRAGM_emphasis	6.6993	2.0434	3.278	0.001044	**
PRAGM_finality	-2.6976	0.6482	-4.162	< 0.001	***
PRAGM_handing over turn	-3.7306	0.4259	-8.758	< 0.001	***
PRAGM_insecurity	-1.5186	1.2638	-1.202	0.229516	
PRAGM_listing	-4.9386	0.4075	-12.118	< 0.001	***
PRAGM_checking	2.3602	1.2075	1.955	0.050632	.
INTERVIEWER_INFLU_post-interaction	6.4369	0.2945	21.858	< 0.001	***
UPTALK_uptalk	13.5252	0.879	15.388	< 0.001	***
Quoted_quoted	8.7541	1.0244	8.545	< 0.001	***
TOPIC_INITIATIONS	-1.2886	0.9069	-1.421	0.156718	
IM_LENGTH	-7.2443	4.3201	-1.677	0.09495	.
REGION_SAmE_HL_LOW	-10.0564	0.7478	-13.449	< 0.001	***
REGION_ AmEO_HL_LOW	-4.758	0.9453	-5.033	< 0.001	***
REGION_CzE_HL_LOW	-5.256	0.699	-7.52	< 0.001	***
REGION_GerE_HL_LOW	-2.326	0.6726	-3.458	< 0.001	***
REGION_MadE_HL_LOW	-4.4234	0.7585	-5.832	< 0.001	***
REGION_MurE_HL_LOW	2.7247	0.8144	3.346	< 0.001	***

Number of obs.=44,153, Random intercepts: Speaker SD (19.90), Residual SD (21.18); Log likelihood =-198033, Conditional $R^{2=}0.83$.

The results are visualized in the effect plots in Figure 31. NBrE female speakers are predicted to produce a mean f0 of 225 Hz for high pitch accents in a continuative function in monologues in pre-interactions, non-uptalking and non-quoted speech with the number of topic initiations and ip length accounted for. All other varieties have a higher f0 for this context but it is only significantly different for SAmE (13.44 Hz more, p=0.006585**), MadE (10.94 Hz more, p=0.035545*), and GerE speakers (13.48 Hz more, p=0.003986**). Low tones are generally 28.21 Hz lower (p<0.001***) than a high tone in the same context. The high tones of male speakers are 79.34 Hz lower (p<0.001***) (Plot 1 in Figure 31). High edge tones are 4.94 Hz lower (p<0.001***) (Plot 2). In the function of emphasis, high tones are generally 6.69 Hz higher (p=0.001044**), finality tunes are 2.69 Hz lower, handing over turn 3.73 Hz lower, and listing by 4.94 Hz lower (all p<0.001***) (Plot 4). The pragmatic functions of checking and insecurity are not significantly different from the continuative function in f0 level by tone (Plot 4). Post-interactions (i.e. utterance by interviewee after interviewer question/statement) are raising the f0 by 6.44 Hz (p<0.001***) (Plot 5). F0 is also higher in uptalk utterances by 13.53 Hz (Plot 6) and 8.75 Hz higher in quoted speech (both p<0.001***) (Plot 7). The number of topic initiations and ip length seem to lower the overall f0 level for high pitch accents but do not exert a significant influence (Plot 8). The interaction REGION*HL shows that all varieties are significantly (all $p<0.001$ ***) different from the NBrE low tones, i.e. with all but the MurE speakers producing lower lows, with the SAmE speakers deviating the most with 10 Hz lower compared to the NBrE lows (Plot 10).

An anova test of the predictor REGION*HL of the final trimmed model was deemed highly significant ($X^2(13)=16218$, $p<0.001$***). In sum, the conditional R^2 values explain 83% of the variance in the data. Additionally, H0 was not rejected for RQ5 for every regional variety for an f0 level analysis by tone label, since GerE and MadE speakers produced significantly higher high tones and almost all varieties produced significantly lower lows compared to the NBrE intercept. These results are in contrast to the f0 level analysis by ip (i.e. ip tune) in Section 4.3.3.

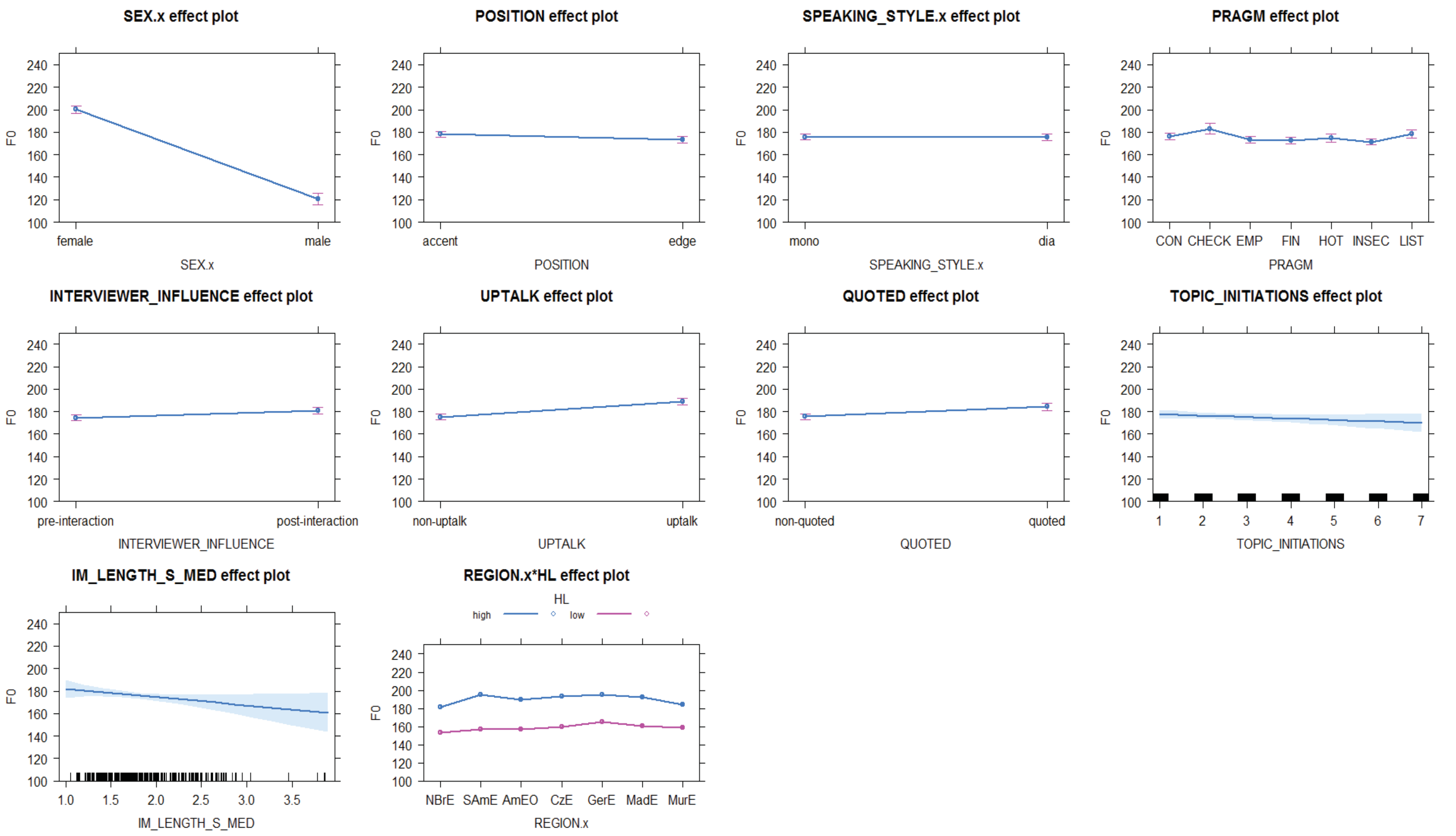


Figure 31. Effect plots for mixed effects model results for f0 by tone label

4.3.2 F0 range by intermediate phrases

Since previous research has found that f0 range is position-sensitive, I will only look at the f0 range by ip for the measures of f0 level and span. Looking at IP measures only would obscure the finer differences between native and non-native speech. Furthermore, based on previous research, gender-based differences will also be analyzed. Thus, linguistic measures (by pattern), by sex (male vs. female), pragmatic functions, speaking style (mono vs. dia), and the position (final vs. non-final) will be considered. F0 level measures are given in Hz and f0 span are given in STs and ERB. First, outliers had to be identified, i.e. f0 values at the extremes (<75 Hz and >350 Hz, and f0 spans that were >200 Hz or negative). After the manual examination, 133 ips and their f0 values (0.71%) out of 18,641 total ips were excluded because they were deemed unreliable. The remaining 18,508 ips and their f0 values served as the basis for the statistical analysis. Quoted and uptalking speech cannot be tested on the basis of a by-ip analysis because these can also occur only in a part of an ip and usually do not stretch over entire ips.

4.3.3 F0 level

The f0 level, i.e. the overall f0 height, was taken as the mean value of the minimum and maximum f0 value of an ip. Table 31 gives an overview of the f0 level in Hz by REGION and SEX.

Table 31. F0 statistics per ip by REGION and SEX: mean f0 levels, minima, maxima, and SD

REGION	SEX	F0_MEAN	SD	F0_MIN	SD	F0_MAX	SD
NBrE	female	195	39.12	167	44.08	224	48.03
SAmE	female	203	42.83	179	36.81	229	60.11
AmEO	female	211	43.27	183	41.76	242	57.45
CzE	female	203	34.59	178	36.99	229	45.16
GerE	female	204	39.67	180	43.59	229	49.84
MadE	female	209	33.71	189	31.76	229	43.03
MurE	female	199	32.62	179	29.97	220	45.25
NBrE	male	114	25.67	99	28.33	130	34.10
SAmE	male	124	24.14	112	20.58	138	31.67
AmEO	male	114	31.99	100	26.43	129	42.57
CzE	male	133	26.15	119	22.09	147	36.25
GerE	male	118	27.95	108	27.80	128	35.63
MadE	male	111	19.82	102	17.69	121	28.56
MurE	male	131	22.03	119	19.03	143	30.64

If the f0 level is taken by ip, the following results can be observed, visualized in boxplots in Figure 69 (see Appendix). The AmEO produced the highest f0 level, while the NBrE females produced the lowest. However, the values are very close and do not diverge drastically. Interestingly, the male NNSs (CzE and MurE) produced the highest f0 level compared to most NSs (except SAmE). One can see in Figure 70 (see Appendix) that in all corpora and both sexes the f0 levels are also position sensitive, i.e. final f0 levels are lower than non-final ones. Position-sensitivity becomes especially evident in female speech, where differences between

15-22 Hz can be found from non-final to final position, whereas 3-16 Hz differences can be found in male speech. Especially the AmEO females show the largest differences between final and non-final f0 levels.

Figure 71 (see Appendix) shows the four main conflated tune patterns by REGION. In female speech LL-tunes are clearly the ones with the lowest f0 level (mean) and HH-tunes the ones with the highest f0 level mean. For the female NSs the difference between HH-tunes is more pronounced than in non-native speech (especially male non-native speech). While for the native females the LH-tune is lower than the most common HL-tune (NBrE) or just as high as the HL-tune (both AmE groups), the LH-tune f0 level is higher in non-native speech, especially those produced by the MurE female speakers. For the male speakers the LL-tune is also the one with the lowest f0 level (except the CzE males). However, the differences between the other tunes are not as pronounced as in the female speakers, especially for the CzE speakers whose f0 level seems to be almost equally distributed in all four tunes.

Similarities between the pragmatic functions across the corpora are that the functions FIN and HOT are generally lower in f0 than CON (see Figure 72 in the Appendix). FIN seems to be lower in the native varieties than HOT. The function of EMP sticks out with the highest f0 level, especially for the native and MadE speakers. The GerE and CzE speakers produce the highest f0 level for CHECK, which is used very infrequently in the other corpora, or not at all by the SAmE and MadE speakers.

The boxplots in Figure 73 (see Appendix) show that in female speech there are barely any differences in speaking style, whereas in male speech there is a tendency for monologues to exhibit a slightly higher f0, except for the German learners for whom both speaking styles have the same level. The final boxplots (see Figure 74 in the Appendix) shows that f0 level is raised in post-interactions for all groups, except the NBrE males where it is almost the same, and for SAmE and the Madrid speakers the post-interactions are slightly lower than the pre-interactions. The descriptive statistics summarized so far, need to be corroborated by inferential statistics by employing a linear mixed effects model to address RQ5. The statistical test addressed the following hypotheses formulated for RQ5 and repeated here for the reader's convenience:

- H5: F0 level measures of NNSs will be closer to the TL norm
- H6: F0 level differences will be position-sensitive
- H7: F0 level will differ by ip length
- H8: F0 level will be higher in monologues
- H9: F0 level will be higher in a checking than in handing over turn
- H10: F0 level will be higher in post-interactions
- H11: F0 level will be higher the more topics are initiated
- H12: F0 level will be higher the more hesitations are produced, because restarts and many smaller ips might contribute to pitch resetting (cf. Hincks 2004)

To verify whether my hypotheses were correct, the following model (conditional $R^{2=}0.86$) was fit: F0_LEVEL_MEAN ~ TUNE_PATTERN_confl * REGION + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + TOPIC_INITIATIONS+ LENGTH + HP_NORM + (1 | SPKR). Residual plots showed deviation from normality and the data was trimmed to exclude data points with residuals deviating by more than two standard deviations. The results of the trimmed model are reported in Table 62 in the Appendix and visualized in Figure 32.

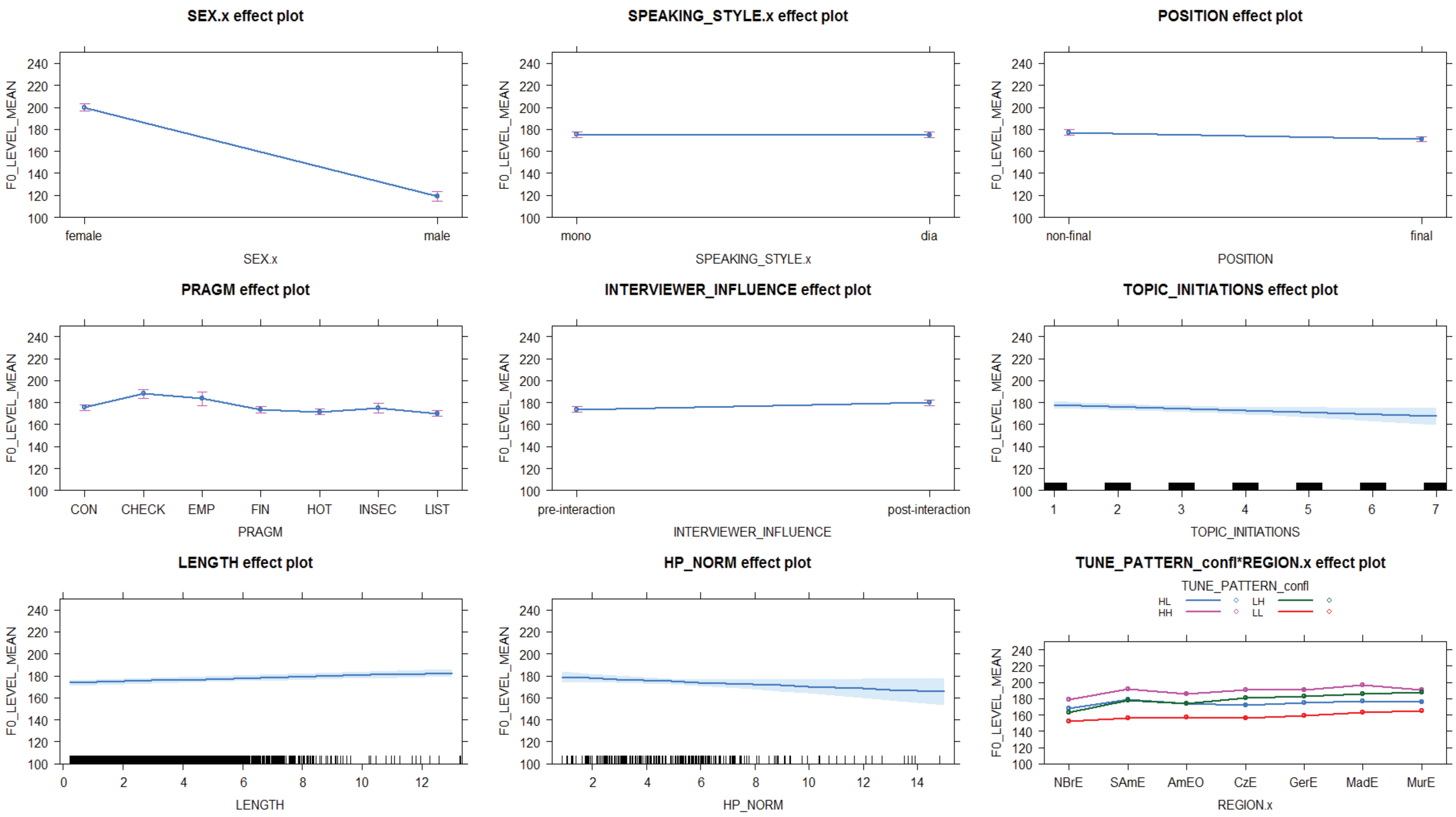


Figure 32. Effect plots for mixed effects model results of f0 level (mean) per ip in Hz

NBrE females in monologues in IP-non-final position in pre-interactions in a continuative pragmatic function with a HL-tune pattern produce an f0 level (mean f0) of 202 Hz, all other factors accounted for. The other three tunes (conflated) are all significantly different from the HL-pattern. The HH-tune (including HH with narrow and wide f0 range), is by 10.81 Hz higher ($p<0.001$***), the LH-tune is by 4.57 Hz lower ($p=0.0125$*), and the LL-tune (including both narrow and wide f0 range LL-tunes) is by 15.99 Hz lower ($p<0.001$***). Compared to NBrE speakers all other varieties have higher f0 levels. Nevertheless, only the SAmE speakers reach significance with 11.36 Hz higher f0 levels ($p=0.0154$*). More significant changes in f0 level can be observed from male to female speech, with male speakers producing ips with 80.75 Hz less ($p<0.001$***) (Plot 1 in Figure 32). Dialogues seem to be produced with a marginally lower f0 level, although the difference is not significant (Plot 2). Another significant effect is ip position; in final positions f0 levels are lower by 6.12 Hz ($p<0.001$***) (Plot 3). In addition to that, five pragmatic functions are significantly different from the CON function, i.e. the CHECK function is by 12.39 Hz higher ($p<0.001$***), the EMP function is by 7.9 Hz higher ($p=0.0063$**), the FIN function is by 1.94 Hz lower ($p=0.039872$*), HOT by 3.96 Hz lower ($p<0.001$***), and the LIST function by 5.42 Hz lower ($p<0.001$***) (Plot 4). In post-interactions, the f0 level is 6.09 Hz higher ($p<0.001$***) (Plot 5). There seems to be a tendency towards lower f0 levels the more different topics are initiated ($\beta=-1.77$ Hz, $p=0.0408$*) (Plot 6). The length of an ip also affects the f0 level, for each additional second the f0 level increases by 0.7 Hz ($p<0.001$***) (Plot 7). The number of hesitations seems to lower the overall f0 level but the difference is not significant (Plot 8). Adding further fluency variables (EL, FP, UP, and WPS) and the AGE variable to the model did not yield any significant results and lead to a lower R^2 value. The interaction between tune pattern (TUNE_PATTERN_confl) and REGION also returned several significant results (Plot 9). The LL-tunes by SAmE speakers are significantly lower by 6.9 Hz ($p<0.001$***). All NNSs groups show the same tendencies compared to the NBrE intercept with highly significant differences, i.e. higher HH-tunes (CzE=7.26 Hz, GerE=5.07 Hz, MadE=8.35 Hz, MurE=4.27 Hz ($p=0.0013$**) and LL-tunes (CzE=12.97 Hz, GerE=12.43 Hz, MadE=12.69 Hz, MurE=17.09 Hz). The MurE speakers are the only ones who in addition to that have 4.91 Hz higher LL-tunes ($p=0.0046$**). An anova test of the null model and trimmed f0 model determined the interaction of TUNE_PATTERN*REGION to be highly significant ($X^2(27)=4764$, $p<0.001$***). Overall the hypothesis for RQ5 that NNSs' f0 level measures will be closer to the TL norm could be confirmed only for the HL-tune, thus corroborating the results of previous f0 level studies. However, the hypotheses were disproved for LH- and HH-tunes, as well as LL-tunes for MurE speakers.

If only the complete learner data is considered in a separate mixed effects model (trimmed to exclude residuals deviating by more than 2 SDs, number of observations: 8763, 115 speakers, conditional $R^2=0.84$, logLik=-38205) with the same predictor variables (plus the learner variables) in order to address RQ11e (*Do learners produce a lower f0 level for LH- and HH-tunes after a stay abroad?*), no differences in f0 level can be observed for the learner varieties and the different tune combinations. None of the learner variables and their interactions (especially REGION*STAB) turned out to be significant. Therefore, the length of the stay abroad and the other learner variables do not contribute to an improvement of LH- and HH-tunes towards the target norm. Overall, the same trends can be observed as for the model above with dialogues, final positions of tunes, and male speakers exhibiting lower f0 levels, and the length of ips and post-interactions raising the f0 level. The same pragmatic effects can also be observed. The only difference to the model above is that the number of hesitations significantly lowers the overall f0 level by 1.6 Hz (p=0.049187*).

4.3.4 F0 span

F0 span measures were transformed to STs, because f0 level transformation resulted in monotone scales which yield the same results in statistical analyses with either Hz or ERB (cf. Mennen et al. 2012: 2255). Previous studies found that f0 span in STs captured cross-language differences marginally better than Hz measures (e.g. Mennen et al. 2012). Descriptive statistics are given in Table 32 in STs, Hz, and ERB; the statistical analysis is, however, based on STs.

Table 32. F0 span mean in STs, Hz, ERB, and SD by REGION and SEX

REGION	SEX	F0 Span (ST)	SD	F0 Span (Hz)	SD	F0 Span (ERB)	SD
NBrE	female	5.38	4.85	57	48.01	1.94	1.45
SAmE	female	4.13	3.81	51	50.34	1.72	1.50
AmEO	female	4.96	4.27	59	51.07	2.02	1.54
CzE	female	4.47	3.91	51	42.96	1.77	1.33
GerE	female	4.25	4.13	48	45.05	1.68	1.40
MadE	female	3.30	2.72	40	33.79	1.43	1.09
MurE	female	3.59	3.20	41	38.70	1.48	1.21
NBrE	male	3.81	3.31	27	28.32	1.02	0.94
SAmE	male	3.56	2.74	26	22.99	0.99	0.80
AmEO	male	3.83	3.12	28	28.40	1.02	0.95
CzE	male	3.58	2.85	28	27.02	1.05	0.90
GerE	male	2.69	2.79	19	25.06	0.73	0.81
MadE	male	3.03	2.89	21	23.83	0.78	0.80
MurE	male	3.08	2.69	24	24.74	0.90	0.82

The speaker groups with the widest spans are NBrE females (5.38 STs) and AmEO females (4.96 STs). While the trend seems to be that the NSs produce a wider f0 span than NNSs, the CzE (4.47 STs) and GerE speakers (4.25 STs) reach native-like values. The female SpE speakers from both regions (MadE=3.30 STs, MurE=3.59 STs), however, seem to behave very similarly, with female and male speakers producing narrower f0 spans. Together with the MadE males (3.03 STs), the GerE males (2.69 STs) produce the narrowest f0 spans of all groups. While the standard deviation seems to be similar in all varieties, the AmEO females (4.27 STs) and NBrE females (4.85 STs) exhibit the widest dispersions, which are almost as high as the mean itself, while the SpE speakers and most male NNSs exhibit the lowest variation in SD values. The same distribution is visualized in Figure 75 (see Appendix). ERB and Hz measures show exactly the same trends as the boxplots in STs in Figure 75 (see Appendix). Even though the ST measure normalizes differences between male and female f0, all other corpora show a tendency of female speakers to exhibit a higher f0 span as compared to male speakers. In female speech, f0 span is wider in final position (see Figure 76 in the Appendix), except for the two SpE corpora where both positions have an equal f0 span. For the male speakers f0 span is also slightly wider in IP-final position, although here the differences are not as pronounced. Figure 77 (see Appendix) shows that in native speech (both male and female) the f0 span is widest in the HL-tune, whereas in non-native speech the LH-tune is either equally wide or wider than the HL-tune. LL- and HH-tunes are produced with the narrowest f0 spans across all varieties. The MadE speakers have almost an equal distribution of f0 span across all pragmatic functions (see Figure 78 in the Appendix). All other learners, who use the checking function, produce the widest f0 span for that function. The listing function has the narrowest f0 span across all varieties. F0 span across speaking styles is extremely similar in both sexes per group (see Figure 79 in the Appendix). The only groups that stick out are the AmEO females with a clearly wider span in dialogues compared to monologues and AmEO males with a narrower f0 span in dialogues compared to monologues produced by female speakers. Figure 80 (see Appendix) shows that there is not much difference in pre- and post-interactions concerning the interaction types. Only AmEO females and CzE males have clearly wider f0 spans (STs) in post-interactions.

If f0 span and f0 level measures are compared together, the following picture emerges (see Figure 33).

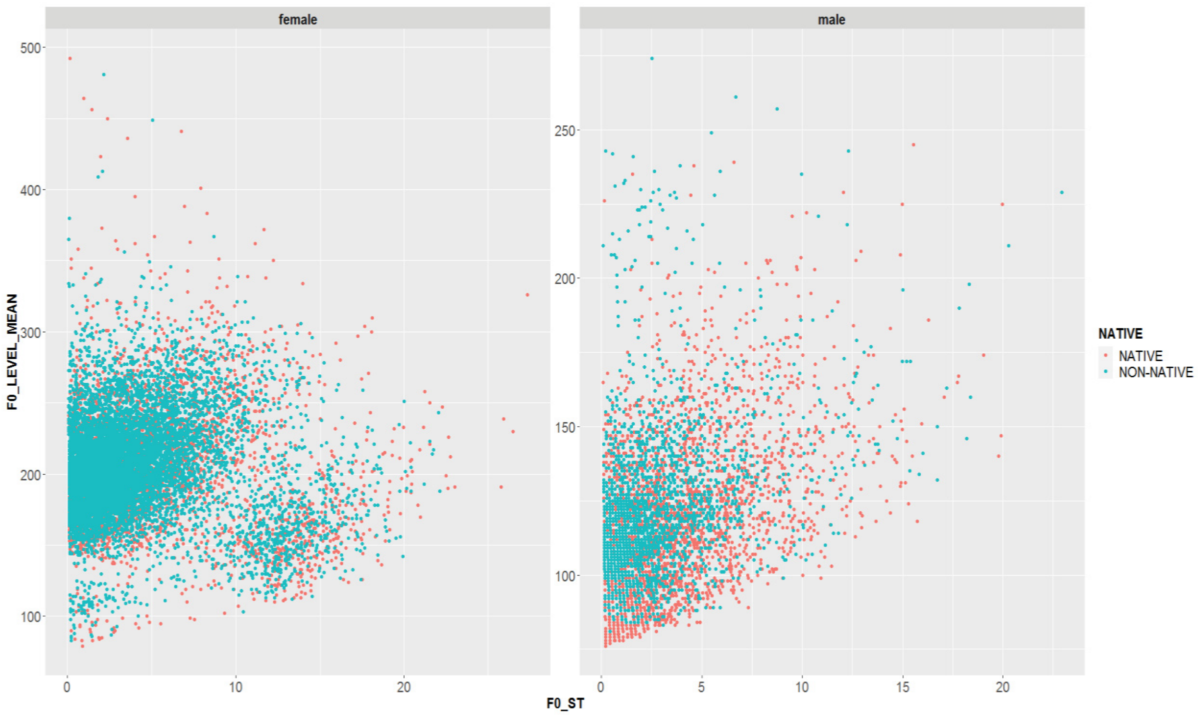


Figure 33. F0 span (in STs) and level (mean in Hz) by native speaker status and SEX

Figure 33 summarizes the f0 span and level and shows an interesting finding, i.e. native and non-native speakers do not differ much in their f0 level and span, since they mostly cluster around the same medians and there are equally frequent outliers in both groups. The medians of the male speakers, however, seem to be more dispersed, while the female speakers seem to almost cluster into two sub-groups, i.e. those females with higher f0 levels and lower spans (most of the data points) and those with lower levels higher spans (fewer data points).

The f0 span and PDQ measures proved to be the exact same measure because of a positive correlation between the two factors of r=0.99. Since they do not measure two different features, no linear regression will be undertaken for the PDQ values. However, for comparisons with previous studies that used PDQ values, the descriptive statistics are reported in the remainder of the present section. Hincks' (2004) formed the hypothesis on the basis of their results that PDQ values around 0.10 characterize "nearly monotonous speech" and those values around 0.25 are typical for "lively and engaging speech" (Hincks 2004: 65). The PDQ values by REGION and SEX are visualized in Table 33 and Figure 34. The medians of all speakers are 0.10 or higher, except for the GerE, MadE, and MurE males, who are below a PDQ-value of 0.10 and, therefore, exhibit the most monotonous speaking styles. Overall, however, the learners and NSs are quite similar in their PDQ values, which was already indicated in the f0 span analysis above.

Table 33. PDQ mean, SD, median, and IQR by REGION and SEX

REGION	SEX	PDQ (mean)	SD	PDQ (median)	IQR
NBrE	female	0.19	0.17	0.13	0.20
SAmE	female	0.15	0.13	0.12	0.16
AmEO	female	0.18	0.15	0.14	0.21
CzE	female	0.15	0.12	0.11	0.14
GerE	female	0.15	0.16	0.10	0.14
MadE	female	0.12	0.09	0.10	0,12
MurE	female	0.13	0.11	0.10	0.12
NBrE	male	0.20	0.30	0.11	0.17
SAmE	male	0.13	0.10	0.10	0.13
AmEO	male	0.15	0.18	0.11	0.15
CzE	male	0.12	0.13	0.10	0.11
GerE	male	0.11	0.15	0.07	0.09
MadE	male	0.12	0.13	0.08	0.11
MurE	male	0.11	0.09	0.09	0.10

Additionally, the results are visualized in Figure 34. This figure shows that native female NSs are more likely to exhibit a more lively speech (i.e. PDQ) and that NNSs, both male and female, are slightly below the native PDQ-values. Nevertheless, the outliers show that for each group there are numerous exceptions with extremely high PDQ-values.

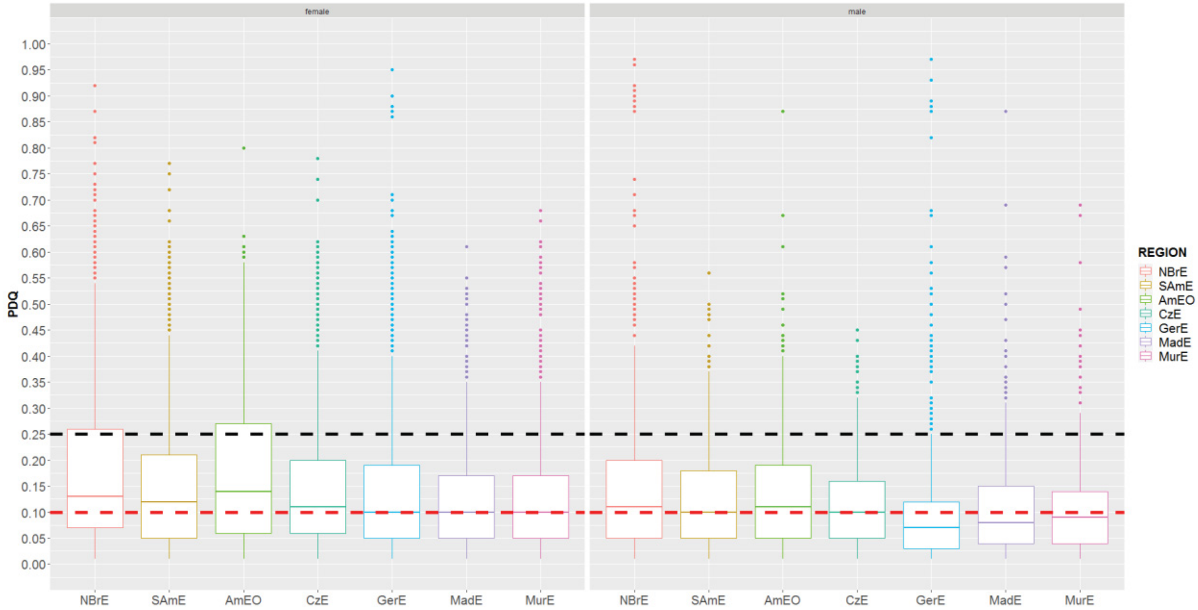


Figure 34. Boxplots for PDQ values (SD of f0 distribution/mean f0 level) by REGION and SEX. Red dashed line=0.10 PDQ=monotonous speech, and 0.25 PDQ=lively speech

The ggplot in Figure 35 shows the PDQ by the four conflated tunes (HL, HH, LH, and LL). The learner varieties in the female section are very similar to each other, with HL- and LH-tunes having the highest PDQ and HH- and LL-tunes the lowest. Generally, this applies to the three NS groups too but the HL-pattern is a lot more "livelier" in female native speech. If the male speakers are considered, the native and non-native speakers are quite similar, especially when it comes to the usage of LL-, LH-, and HH-tunes. However, the HL-tunes have a higher PDQ in male native speech than in male non-native speech.

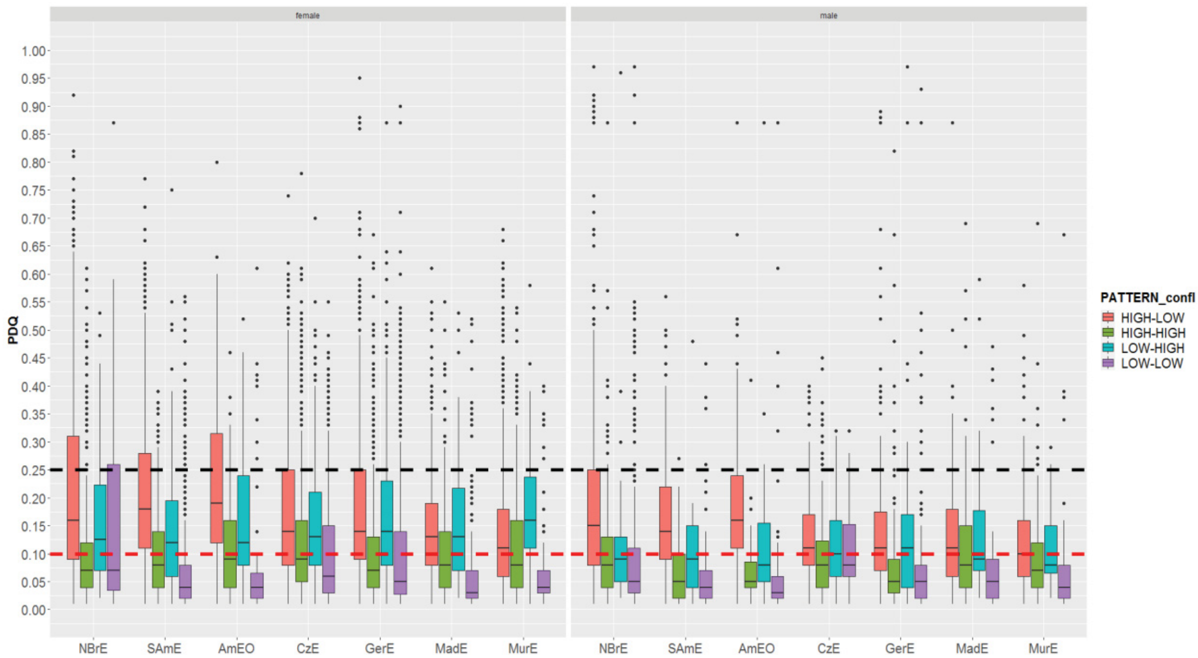


Figure 35. Boxplots for PDQ values (SD of f0 distribution/mean f0 level) by REGION, SEX, and ip tune. Red dashed line=0.10 PDQ=monotonous speech, and 0.25 PDQ=lively speech

The plots in Figure 34 and 35 show that there are many outliers in the PDQ-values. Future studies would have to analyze individual performances of speakers more closely.

To get back to the f0 span analysis, the descriptive statistics are complemented by an inferential analysis (based on f0 span measured in STs), which addresses the following research question: RQ6: *Do learners of English produce a narrower f0 span than NSs? Does this differ depending on the speaking style, tune pattern, sex, number of topic initiations, position in the IP, ip length, and pragmatic function?*

H0 the f0 span of the NNSs is not narrower than that of the NSs, and does not differ depending on the levels of the IVs: SPEAKING_STYLE, TUNE_PATTERN_confl, SEX, POSITION, LENGTH, TOPIC_INITIATIONS, PRAGM: $R^2 = 0$.

H1 the f0 span of the NNSs is narrower than that of the NSs, and differs depending on the levels of the IVs: SPEAKING_STYLE, TUNE_PATTERN_confl, SEX, POSITION, LENGTH, TOPIC_INITIATIONS, PRAGM: $R^2 > 0$.

The following directional hypotheses were formulated for RQ6, repeated here for the reader's convenience:

H13: F0 span measures of NNSs will be narrower than those of NSs
H14: F0 span differences will be position-sensitive
H15: F0 span will be wider in longer ips
H16: F0 span will be wider in dialogues
H17: F0 span will be wider in checking than in handing over turn
H18: F0 span will be wider in post-interactions
H19: F0 span will be wider the more topics are initiated
H20: F0 span will be wider the more hesitations are produced

In order to addresses the overall research question and the directional hypotheses, the following model (conditional $R^{2=}0.38$) was fit: F0_SPAN_ST ~ TUNE_PATTERN_confl * REGION + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + LENGTH + TOPIC_INITIATIONS + (1 | SPKR). The residual plots showed some deviation from normality and thus the model was trimmed to exclude data points with residuals deviating by more than 1.5 standard deviations. The results are visualized in the effect plots in Figure 36.

The mixed-effects model (see Table 63 in the Appendix for the results) revealed that NBrE females in monologues in IP-non-final position in pre-interactions in a continuative pragmatic function with a HL-tune are predicted to produce an f0 span of 5.05 STs. All other tunes have a narrower span in NBrE, i.e. HH - 2.72 STs, LH -1.49 STs, and LL -2.27 STs. The AmE speakers do not have a significantly different f0 span from the NBrE intercept. All learner groups have a narrower span, i.e. CzE (-0.68 STs, $p=0.01226$*), GerE (-0.82 STs, $p=0.00126$**), MadE (-1.81 STs, $p<0.001$***), and MurE (-1.91 STs, $p<0.001$***). Even though the conversion to STs normalizes most of the differences between male and female f0, male speakers still produce a narrower f0 span (-1.08 STs, $p<0.001$***) (Plot 1 in Figure 36). F0 span is slightly wider in dialogues (0.18 STs, $p=0.02483$*) (Plot 2). IP-final ips have a wider f0 span by 0.78 STs ($p<0.001$***) (Plot 3).

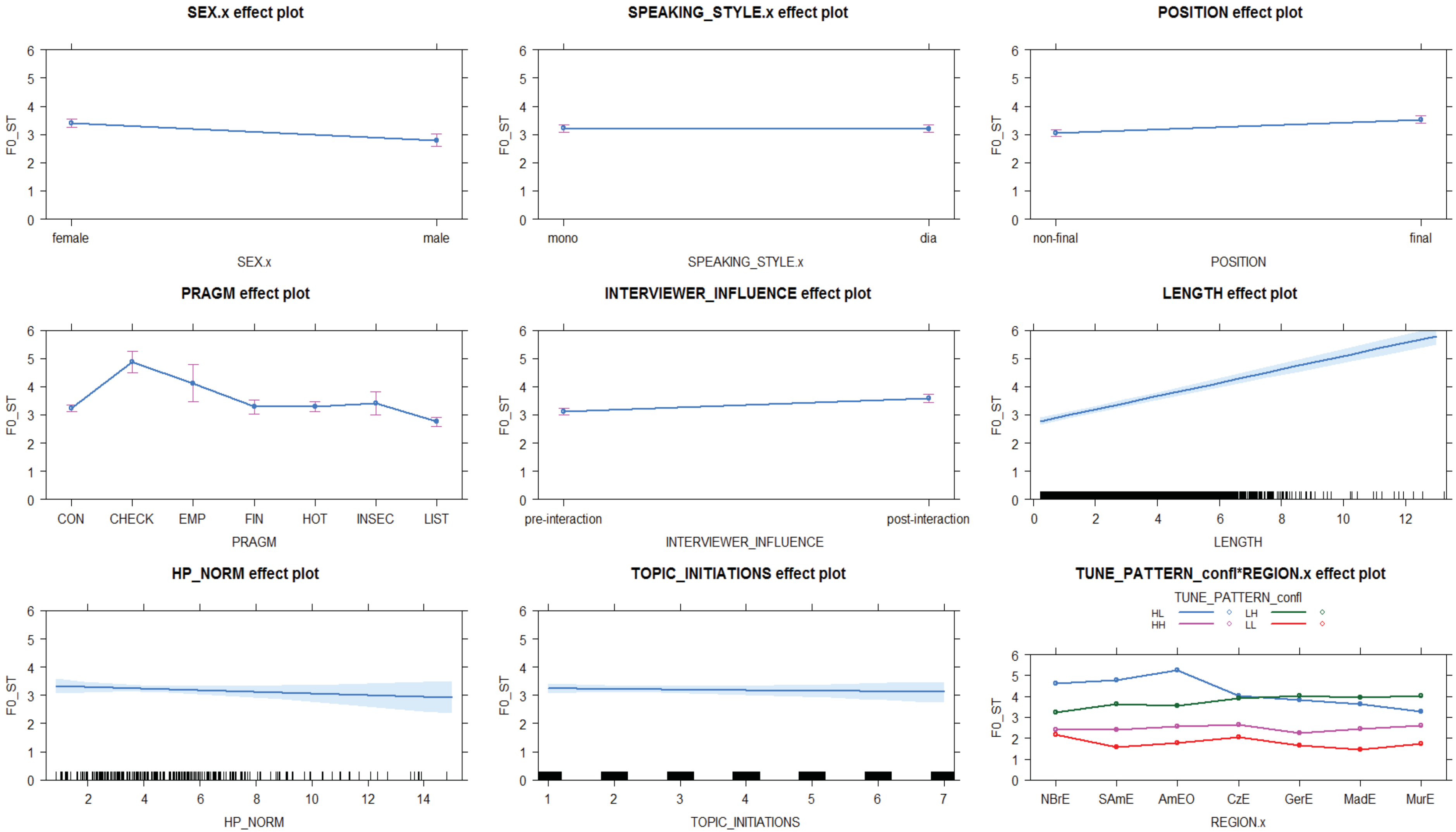


Figure 36. Effect plots for mixed effects model results of f0 span per ip in STs

Also some pragmatic functions show significant effects, i.e. the CHECK function is by 1.12 STs wider ($p<0.001$***), EMP is also wider by 1.12 STs ($p=0.02664$*), HOT is by 0.43 STs wider ($p<0.001$***), and the LIST function is by 0.52 STs narrower ($p<0.001$***) (Plot 4). Post-interactions are by 0.45 STs wider and longer ips are by 0.29 STs wider for every increase in LENGTH (absolute number in seconds per ip) (both $p<0.001$***) (Plots 5 and 6). The number of HPs and topic initiations do not have any significant effects on f0 span (Plots 7 and 8). The interaction REGION*TUNE_PATTERN also showed several significant results ($X^2(27)=3730$, $p<0.001$***) (Plot 9). Both AmE groups have narrower LL-tunes (SAmE: -1.06; AmEO: -1.29, both $p<0.001$***). As with the f0 level results, the learners show striking similarities in f0 span differences by tune pattern. In all patterns except the standard HL-tune, the learners have a wider f0 span, especially in the LH-tune. The LL-tunes, even though slightly wider in f0 span, are not significantly different except for the MurE learners ($\beta=0.75$ STs, $p=0.01922$*). The LH-tunes are wider by 0.96 STs in the CzE learners ($p=0.01124$*), 1.30 STs in the GerE learners, 1.96 STs in the MadE learners, and 1.89 STs in the MurE learners (all are highly significant $p<0.001$***). In addition, the HH-tunes are wider in the learner ips, 0.86 STs Czech, 0.57 STs German ($p=0.00705$**), 1.56 STs Madrid, and 1.93 STs Murcia (all are highly significant $p<0.001$***). Adding further fluency variables (EL, FP, UP, and WPS) and the AGE variable to the model did not yield any significant results and lead to a lower R^2 value. In sum, H1 for RQ6 was only confirmed for HL-tunes for all NNSs, who all produced narrower f0 spans for the HL-tune. For all other tunes, no differences or wider f0 spans were found, confirming one of the hypotheses that f0 span is position-sensitive, since most high-register tunes appear towards the end of IPs.

An additional analysis with the complete learner data and learner variables was conducted in order to answer RQ11f (*Do learners produce a wider f0 span for HL-tunes and narrower span for LH- and HH-tunes after a stay abroad?*). The residual plots of the mixed effects model showed some deviation from normality and thus the DV F0_SPAN_ST was logged (log2). The final model (number of observations: 9,034, 115 speakers, conditional $R^2=0.23$, logLik=-15840) included the following variables: log2(F0_SPAN_ST) ~ TUNE_PATTERN*REGION*STAB + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + LENGTH + TOPIC_INITIATIONS + YOEAS + YOEAU + NO_OFL + (1 | SPKR). The COUNTRY variable was left out because it did not improve model fit. The results are shown in Table 64 in the Appendix. Main effects for SEX ($X^2(16.5)=1$, $p<0.001$***), POSITION ($X^2(254)=1$, $p<0.001$***), PRAGM ($X^2(57.1)=6$, $p<0.001$***), INTERVIEWER_INFLUENCE ($X^2(20.9)=1$, $p<0.001$***), LENGTH ($X^2(136)=1$, $p<0.001$***), and TUNE_PATTERN*REGION*STAB ($X^2(1094)=35$, $p<0.001$***) were found. The main effects are visualized in Figure 37. The analysis determined that males tend to produce a narrower f0 span for HL-tunes overall ($p<0.001$***) (Plot 1 in Figure 37). HL-tunes in final position are significantly wider in their f0 span ($p<0.001$***) (Plot 2). HL-tunes used in CHECK are significantly wider in their f0 ($p<0.001$***), while they are significantly narrower in LIST ($p<0.001$***) (Plot 3). In post-interactions and longer ips in seconds (both $p<0.001$***) (Plots 4 and 5), f0 span becomes wider. GerE speakers who have been on short stays abroad produce wider f0 spans for LH-tunes ($p=0.02957$*) (Plot 6). All other variables were not significant. In sum, the only effect the STAB variable has was for the LH-tune for GerE learners with short stays abroad, who produce wider f0 spans compared to the CzE intercept.

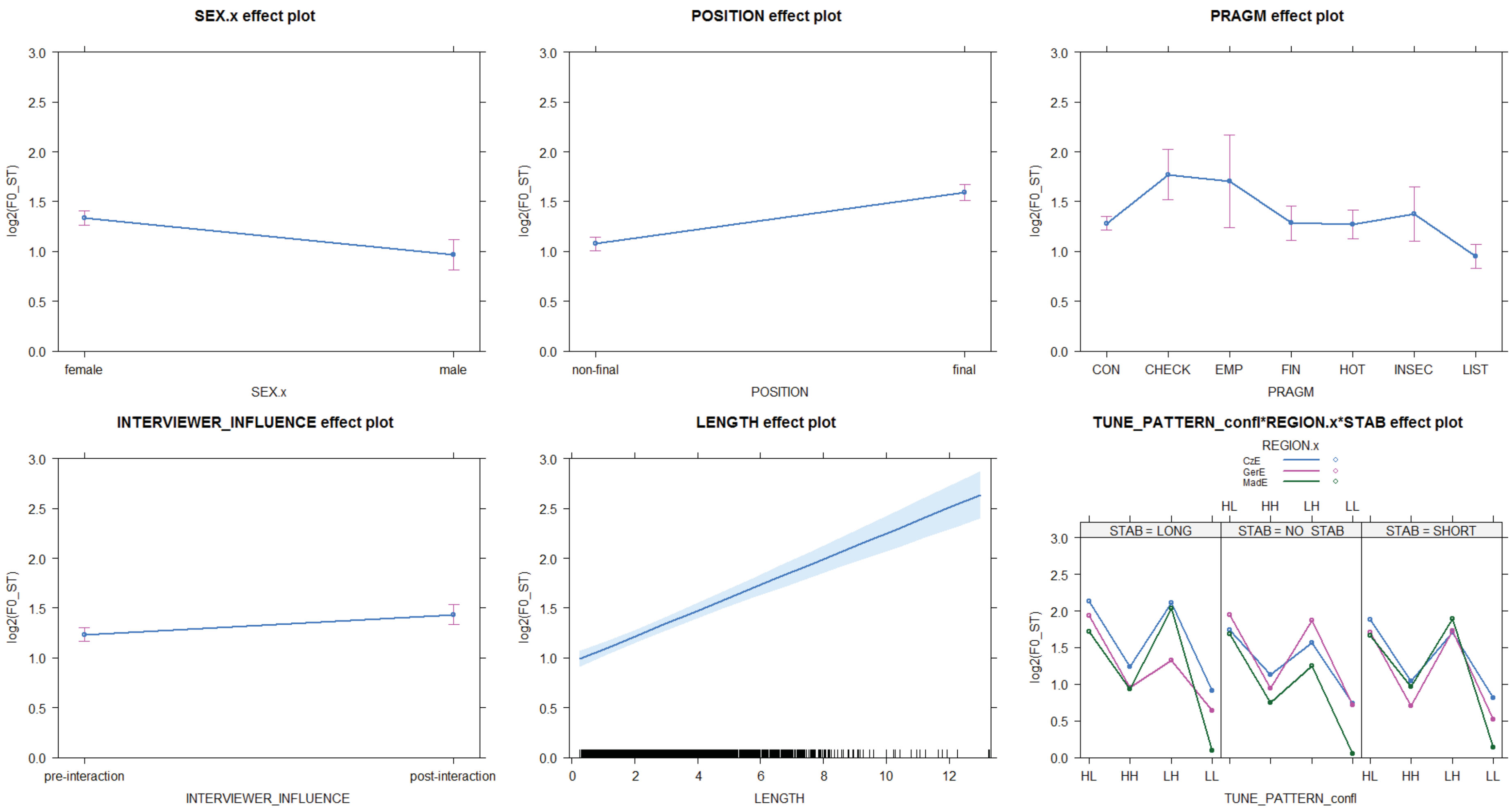


Figure 37. Effect plots for mixed effects model results of f0 span per ip in STs for the complete learner data only (main effects only)

4.3.5 F0 range and entrainment (interviewers vs. interviewees)

F0 range was also investigated in conversations where an interviewer changes his/her f0 range in response to an interviewer or vice versa. The tone label analyses in 4.1.2 and tune analysis in 4.1.6 determined that high tones are more frequent in post-interactions and both tone choice and tune choice seem to be influenced by the interviewer's speech. The previous f0 analyses (see Sections 4.3.1-4.3.5) of f0 range (level and span) included the independent variable INTERVIEWER_INFLUENCE (pre- vs. post-interactions), and determined the following results (see Table 34 and 35):

Table 34. F0 level and INTERVIEWER_INFLUENCE

Fixed effects	**Estimates**	**Standard error**	**t-value**	**p-value**	
INTERVIEWER_INFLUENCE_post-interaction	6.0978	0.3954	15.42	<0.001	***

Table 35. F0 span and INTERVIEWER_INFLUENCE

Fixed effects	**Estimates**	**Standard error**	**t-value**	**p-value**	
INTERVIEWER_INFLUENCE_post-interaction	0.454	0.0704	6.45	<0.001	***

In post-interactions, the f0 level rises by six Hz and by 0.45 STs for f0 span. While this is a quite general analysis by ip, i.e. that global entrainment is taking place, a more detailed local analysis has to be undertaken as well. Levitan and Hirschberg (2011) determined that local entrainment was stronger than global entrainment, and therefore, stronger effects can be expected in the present data set.

In order to analyze entrainment locally, a difference in distance score (DID) was calculated for all conflated tone labels (%H, H*, H-, H%, L*, L-, and L%) per speaker. This measure normalizes f0 by taking into account the f0 level in pre- and post interactions and in relation to the interviewer's f0 by speaker. This measure was obtained by subtracting the absolute difference of interviewer f0 in post-interactions and interviewer f0 from the absolute baseline difference (interviewee f0 in pre-interactions - interviewer f0) for each tone label (cf. Pardo et al. 2013; Lewandowski and Nygaard 2018). The resulting DID-scores then indicate whether the distance between interviewer speech and interviewer in pre- and post-interactions is smaller or larger than the baseline distance. A positive value is indicative for entrainment, while a negative value shows disentrainment, and a value of zero shows no change, i.e. maintenance of f0 level. The larger the values are, in either direction, the larger or smaller the extent of entrainment/disentrainment can be observed (cf. Lewandowski and Nygaard 2018).

The analysis is based on 159 speakers, since some speakers include purely monologic speech without any interviewer interventions (at least in the selected part for this study), while others include only one to two interviewer interventions, or others have no monologic part at all. Furthermore, speakers who produce only answers or only statements were excluded as well, since no reliable comparison can be made between the two speech acts. Thus, 66 speakers out of 225 were excluded from the analysis in the present section.

DID-scores reached from -83 Hz to 116 Hz with a mean of 8.22 Hz (21.1 SD; median=7 Hz, IQR=21). A decrease of distance (indicated by positive values) indicates entrainment and an increase (indicated by negative values) indicates disentrainment. The mean shows that on average the interlocutors entrain to each other. Figure 38 shows that most DID-scores for all tone labels are positive. Generally, earlier parts of an utterance, i.e. H% or H*, have the highest DID-scores in most of the varieties of English investigated. The MurE females deviate the most with unusually low %H- tones, indicating that MurE female speakers may disentrain in post-interactions. The SAmE females seem to maintain their f0 level in earlier parts of their utterances immediately following an interviewer statement or question. Interestingly, high boundary tones in the GerE female speakers seems to be positively entrained, almost as much as the initial boundary tones (%H), which could provide further insights into findings reported in earlier section, i.e. that female GerE speakers produce more and higher (f0 level) high boundary tones. It is possible that for GerE females the entrainment effect lasts longer until later parts of an utterance, as compared to other speakers from different regional backgrounds. While on average the NBrE maintain their f0 level for high boundary tones (H%) in post interactions, 25% of the data, nevertheless, overlap with the GerE values, which indicates that for some NBrE female speakers a longer-lasting effect may be found. The MurE males, on the other hand, have high %H-tone DID-scores and the CzE males the highest of all. The SAmE males also show clear early entrainment effects for high pitch accents (H*). The NBrE and GerE males, however, show only slight entrainment effects on all tone labels.

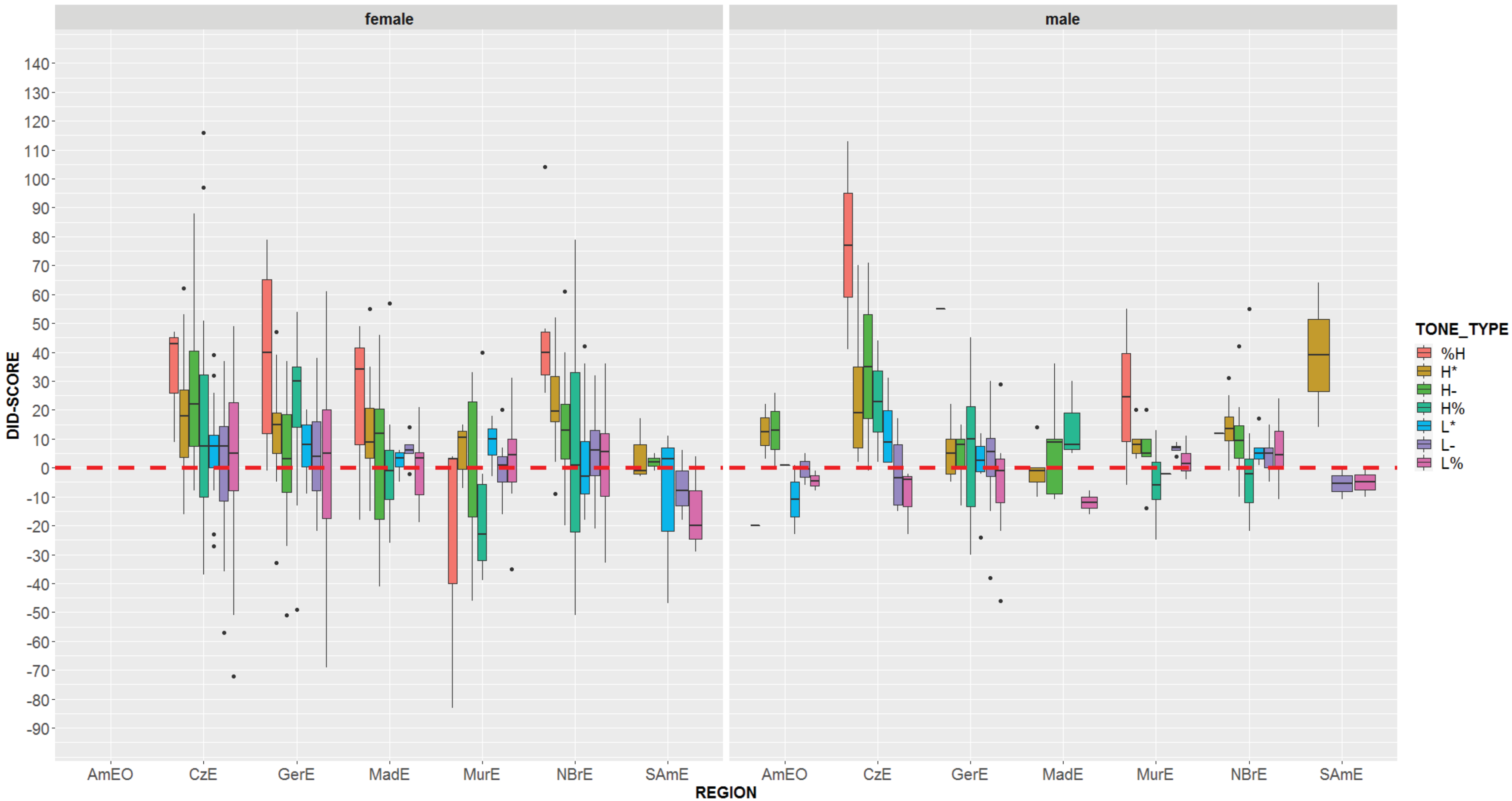


Figure 38. Boxplots for DID-scores by REGION and Tone Label

A mixed-effects model was built to assess entrainment on f0 level by tone labels and to answer RQ7: *Do N(N)Ss entrain their f0 range to that of their respective interviewers?*

H0 the DID-scores of NSs and NNSs are not positive and do not differ depending on the levels of the IVs: TONE_TYPE, REGION, SEX, INT_SEX, and TIME_IN_RECORDING: $R^2 = 0$.

H1 the DID-scores of NSs and NNSs are positive and differ depending on the levels of the IVs: TONE_TYPE, REGION, SEX, INT_SEX, and TIME_IN_RECORDING: $R^2 > 0$.

The f0 DID-score was introduced to the model (conditional $R^{2=}0.24$) as the DV. The main fixed effect of interest was the tone label type (TONE_TYPE) and the position of the tone to determine position-sensitive entrainment and local entrainment. REGION was entered as an IV to uncover differences between NSs and NNSs. An interaction term for SEX and INT_SEX was entered as well, to see whether certain gender-pairings lead to different entrainment effects. TIME_IN_RECORDING (early vs. late) was entered into the model as another IV, because entrainment effects have been found to increase over time in interactions (cf. Pardo 2006, 2010; Levitan and Hirschberg 2011). However, also early entrainment effects have been found by Pardo (2006, 2010). SPEAKER was added as a random intercept. Visual inspection of residual plots showed slight deviation from normality and, therefore, the model was trimmed to exclude data points with residuals deviating by more than two standard deviations. The results of the mixed effects model are reported in Table 36.

Table 36. Results for mixed effects model with DID scores by tone label. Final model: DID_SCORE ~ TONE_TYPE + REGION + SEX * INT_SEX + TIME_IN_RECORDING + (1|SPKR)

Fixed effects:	**Estimate**	**Std. Error**	**t value**	**p-value**	
(Intercept)	27.992	3.497	8.01	< 0.001	***
TONE_TYPE_H*	-12.557	3.374	-3.72	< 0.001	***
TONE_TYPE_H-	-14.925	3.393	-4.4	< 0.001	***
TONE_TYPE_H%	-19.903	3.536	-5.63	< 0.001	***
TONE_TYPE_L*	-21.62	3.503	-6.17	< 0.001	***
TONE_TYPE_L-	-22.574	3.413	-6.61	< 0.001	***
TONE_TYPE_L%	-22.975	3.418	-6.72	< 0.001	***
REGION_AmEO	0.681	3.089	0.22	0.82578	
REGION_SAmE	-1.191	2.776	-0.43	0.66845	
REGION_CzE	-0.425	2.38	-0.18	0.8586	
REGION_GerE	-3.032	1.904	-1.59	0.11313	
REGION_MadE	-5.122	2.416	-2.12	0.03513	*
REGION_MurE	-6.274	2.311	-2.72	0.00734	**
SEX_male	-4.541	1.528	-2.97	0.00337	**
INT_SEX_male	-1.512	2.666	-0.57	0.57148	
TIME_IN_RECORDING _late	3.39	1.37	2.47	0.01431	*
SEX_male:INT_SEX_male	11.377	5.56	2.05	0.04253	*

Number of obs.=864, Speakers=191, Random intercepts: Speaker SD (5.41), Residual SD (14.40); Log likelihood =-3.577, Conditional $R^{2=}0.24$.

The results are visualized in the effect plots in Figure 39. The mixed effects model's intercept (NBrE females producing %H-tones when the interviewer is female in earlier parts of a recording) was significant and positive, and thus showing overall entrainment in f0 on a %H-tone. All other tone labels are significantly lower in their DID-scores, which suggests that the distance in the interviewers' and interviewees' utterances become gradually higher, distances being the lowest immediately at turn-transitions, i.e. for %H-tones and highest in IP-final tones (i.e. L%) (all differences p<0.001***) (Plot 1 in Figure 39). The AmEO speakers are predicted to exhibit the highest DID-scores (β=0.68, difference to intercept n.s.), while all other groups are predicted to produce lower scores at immediate turn-transition points, the difference only being significant for SpE speakers (MadE: β=-5.12, p=0.03513*; MurE: β=-6.27, p=0.00734**) (Plot 2). In general, male speakers entrain less by -4.54 Hz (p=0.00337**). The time in the interaction (early vs. late) showed that in later parts of a recording the entrainment is larger by 3.39 Hz (p=0.01431*) (Plot 3). The interaction between interviewee and interviewer sex is significant and there seems to be a tendency for male-male interactions to result in higher DID-scores (β=11.38, p=0.04253*) (Plot 4). Model comparisons with anova tests determined that the tone position (initial, pitch accent, phrase accent, and boundary tone) is the most important factor in determining DID-scores and thus f0 entrainment (X^2 (96.9), df=6, p<0.001***). The interaction SPKR_SEX*INT_SEX also proved to be a significant factor (X^2 (10.3), df=3, p=0.016*). TIME_IN_RECORDING was also significant (X^2 (6.06), df=1, p=0.014*). The factor REGION was non-significant, indicating that there are only small differences between native and non-native speech in DID-scores. The interviewers' regional varieties did not improve model fit, and did not show significant differences and were, therefore, left out from the final mixed effects model. This confirms results by previous research which also found that f0 entrainment varied by interviewer but not by the interviewers' accent (cf. Lewandowski and Nygaard 2018).

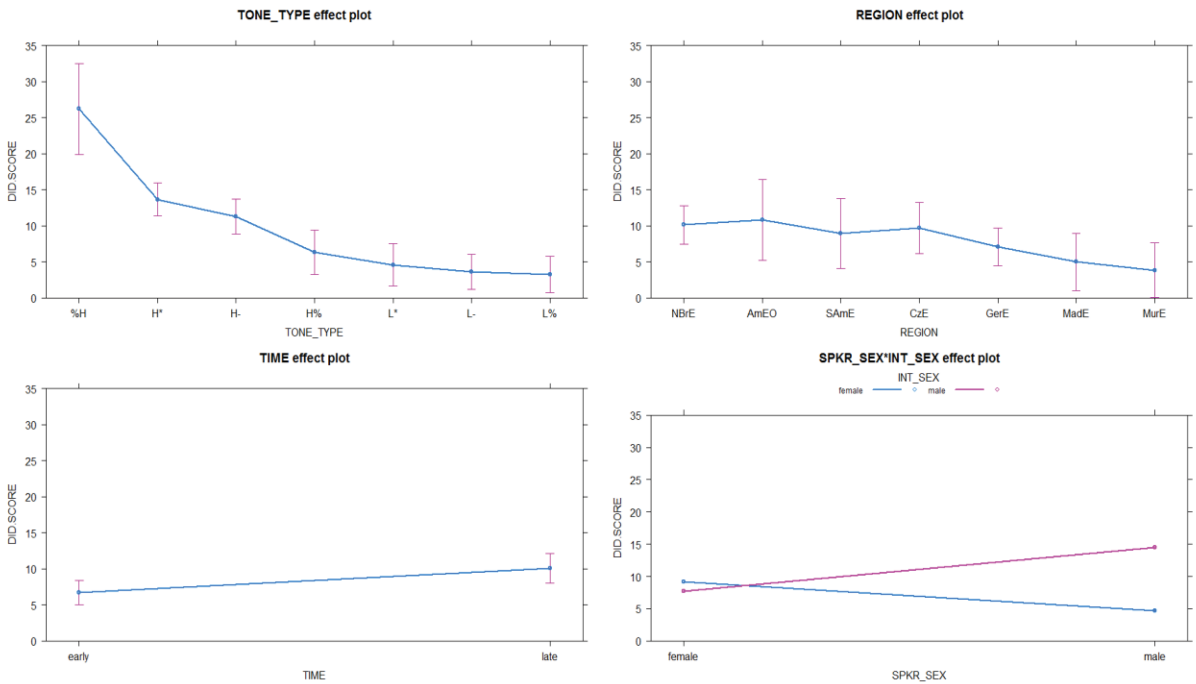


Figure 39. Effect plots for mixed effects model results of DID-scores

Bi-directional phonetic entrainment is possible in my data set. However, it could not be analyzed further because no comparable data from the interviews in L1-L1 or L1-L2 environments was available. All in all, it could be determined that learners match the interviewer's f0 level at the beginning of turns. The overall effect of entrainment seems to be stronger early on in an utterance, in same-gender interactions, and in later parts of a recording. Male speakers were found to entrain less than female ones, confirming the hypothesis that female speakers will entrain more to their interlocutors. Nevertheless, it has to be kept in mind that male-male interactions seem to significantly raise DID-scores. However, the hypothesis that NNSs entrain more than NSs, could not be confirmed. The CzE and GerE speakers were found to produce similar entrainment effects to the NSs, while the SpE speakers entrained significantly less. The MurE speakers were the group who entrained the least compared to the other NNSs. Thus, other explanations may apply to these findings, which will be discussed in Section 5.4. In sum, H0 for RQ7 was rejected, because high DID-scores were found for all varieties of English and thus entrainment effects were found.

4.4 Uptalk in native and non-native speech

The focus of this sub-section is to analyze differences in phonetic realization of uptalk tunes. Since the previous analyses in the present study have shown that learners produce more high edge tones than NSs, the present section takes a closer look at how phonetically different or similar the learners are from each other with regard to the high-register tunes that were labeled as uptalk tunes. In addition, I will zoom in on the pragmatic functions of these extremely high tunes. Basic descriptive statistics are given in Table 37 and Figure 40. In total there are 83 speakers (68 female and 15 male) in the corpus sample of 225 speakers, who at some point in the 560 word samples use uptalk. The age range of the 'uptalkers' is 19-33 years. 141 of the uptalk utterances come from monologues and 180 from the dialogues. Uptalk seems to function differently in the monologues than the dialogues. While in the monologues uptalk seems to look and function differently, i.e. more meaning something like "are you still paying attention to what I am saying?", in the dialogues it tends to indicate something like "do you understand what I am saying?". Table 37 shows that in total 889 of all tones (n=47,442) are used in uptalk tunes (1.87% of all tones). 16% of all upstepped high phrase accents (^H-) and 33.67% are upstepped high boundary tones (^H%), while 14.71% of all L*+H pitch accents are part of uptalk tunes. The rest of the ^H-, ^H%, and L*+H tones are associated with simple continuation rises, with no intention of checking interviewer comprehension or involving the interviewer in the conversation.

Table 37. Proportion of uptalk and non-uptalk tones

LABEL	non-uptalk	uptalk	total_no_tones	% of uptalk tones
!H-	1,129	3	1,132	0.27
!H%	280	0	280	0.00
!H*	225	2	227	0.88
%H	237	0	237	0.00
^H-	708	137	845	16.21
^H%	325	165	490	33.67
^H*	743	13	756	1.72
H-	4,713	137	4,850	2.82
H%	1,096	14	1,110	1.26
H*	14,704	145	14,849	0.98
H+L*	149	1	150	0.67
L-	11,765	49	11,814	0.41
L%	4,744	0	4,744	0.00
L*	5,678	218	5,896	3.70
L*+H	29	5	34	14.71
L+H*	28	0	28	0.00
TOTAL	46,553	889	47,442	1.87

In total there are 321 uptalk tunes, including 889 individual tones. The GerE speakers produce a total of 170 uptalk tunes, the CzE speakers 97 tunes, the MadE speakers 13 tunes, MurE 22 tunes, AmE speakers 17 tunes, and NBrE 2 tunes. Figure 40 shows the uptalk instances proportionally to the number of all ips by speaker and how many of the total number of ips contained uptalk instances.

The uptalk instances are not normally distributed across speakers, with some speakers contributing 1 and others up to 16 instances per 560 words. Additionally, most of the uptalkers are female as can be gleaned from Table 38.

Table 38. Overview of uptalk instances by REGION and SEX

REGION	n	SEX	Uptalk instances
AmE	8 out of 45	1 male, 7 female	1-4
NBrE	2 out of 45	1 male, 1 female	1 each
CzE	26 out of 45	5 male, 21 female	1-15
GerE	32 out of 45	8 male, 24 female	1-16
MadE	7 out of 45	all female	1-3
MurE	6 out of 45	all female	1-7

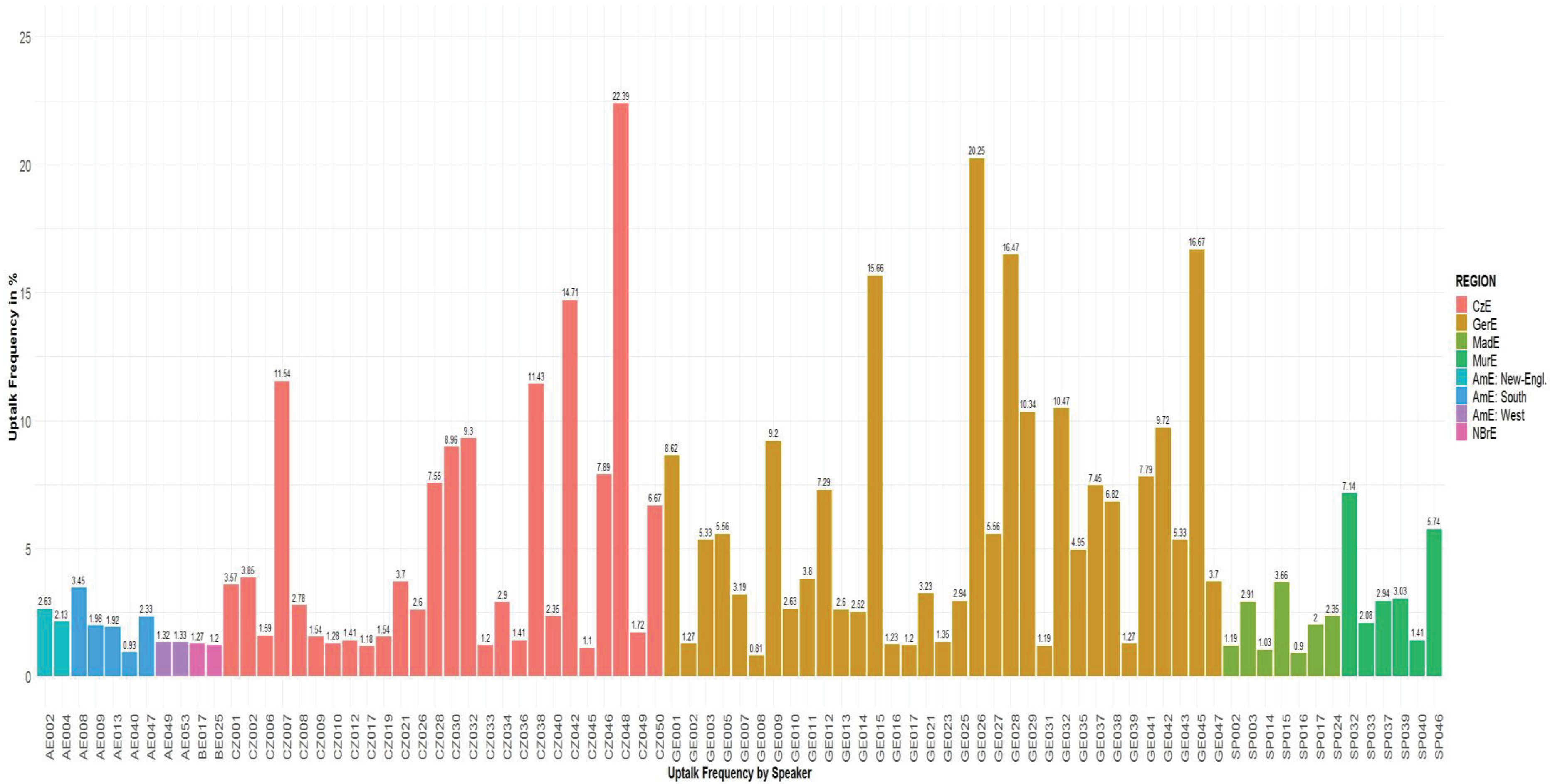


Figure 40. Barplot for uptalk instances (in percent) proportionally to the number of all ips by speaker and REGION

Overall, there is a higher incidence of uptalk tunes in the GerE sample. In the SpE components none of the uptalkers are male. The uptalk tunes are used mainly with four pragmatic functions, i.e. continuation, checking, insecurity, and listing. Examples of each pragmatic function from the corpus samples are given in examples 4-7 with the underlined words representing the position of the uptalking tunes.

(4) Continuation: GE015: "well I er twice for one week I went to a farm because there is this organisation where you can go to a farm and work there for wha=whatever time you want to stay there"

(5) Checking: CZ038: "because I believe that it er deepened my er like understanding of multicultural multiculturality"

(6) Insecurity: CZ038: "and er I was in the er national representation"

(7) Listing: GE015: "because we didn't have we went to a youth hostel and um they didn't have air-condition and it was very er w=wet there"

Example (4) includes speaker GE015 who uses uptalk even after this utterance to signal continuation, i.e. reduplicates the pattern to signal the wish to continue speaking. Example (5) includes an example of an utterance signaling checking. Speaker CZ038 is checking whether the word has been understood, especially since she had a false start with "multicultural". In example (6) speaker CZ038 signals insecurity by the use of uptalk on "representation", where she does not seem sure about the terminology. In example (7) speaker GE015 uses uptalk after every ip to list all the reasons why her experience abroad was not a pleasant one.

On average (median) similar pause lengths (in seconds) are produced by the speakers for each pragmatic function, i.e. (checking 0.43, continuation 0.47, insecurity 0.43, and listing 0.22). For listing items, the tune may make it clear that the speaker intends to hold the floor and, therefore, shorter pauses may be allowed. The speakers that use uptalk especially often are four CzE and four GerE speakers. The profiles of these learners will be inspected more closely (see Table 39). These eight learners are all female, between 21-24 years old, have been abroad, most of them talk about a country (one of the LINDSEI topics in the monologues) (except CZ048). CZ007 was vaguely familiar with the interviewer, and the other three Czech learners were familiar with the male Czech interviewer. How familiar the GerE speakers were with the interviewers is unknown. According to CEFR ratings, CZ007 and GE028 have a B2 level, the other three CzE speakers have a C1 level. The CEFR levels for the three other GerE speakers are unknown at the time of compilation of the present thesis. Interestingly, speaker CZ048 uses 14 out of 15 uptalk instances for a checking function while speaking to a male non-native interviewer. Speaker GE015 uses five out of 13 uptalk instances to signal checking while speaking to a male native AusE speaker. Whether it is a coincidence or a statistically significant result cannot be studied in the present study because the interviewees were only interviewed by one speaker and a comparison cannot be made how the same learners would have behaved if they had been interviewed by a female interviewer. This would have to be investigated in future studies whether speakers use uptalk for different pragmatic functions in mixed-sex or single-sex interactions.

Table 39. Frequent uptalk users in the selected corpus sample

SPKR	**Uptalk instances**	**Sex**	**Age**	**STAB months**	**Country STAB**	**Topic**	**Interviewer Variety**	**INT_SEX**
CZ007	11.54%	f	24	5	Ireland	country	BrE	f
CZ038	11.43%	f	23	28	Ireland	country	CzE	m
CZ042	14.71%	f	21	1.75	GB	country	CzE	m
CZ048	22.39%	f	21	0.75	GB	movie	CzE	m
GE015	15.66%	f	23	6	Australia	country	AusE	m
GE026	20.25%	f	24	3	USA	country	AmE	f
GE028	16.47%	f	23	4	USA	country	AmE	f
GE045	16.67%	f	22	11	GB	country	BrE	f

Figures 41 and 42 show the most frequent uptalk uses in the corpus sample of the present study, i.e. those speakers who produce more than three uptalk instances in the 560 word samples (Table 65 in the Appendix provides further details). In total, there are 19 GerE speakers (two male speakers excluded from overview and subsequent analyses) who use more than three uptalk instances, nine CzE speakers, and two MurE speakers (these were left out of the overview and subsequent analyses). There is a tendency for these speakers to use a LH- or HH-tune to fulfill a continuative and checking function (see Figures 41 and 42).

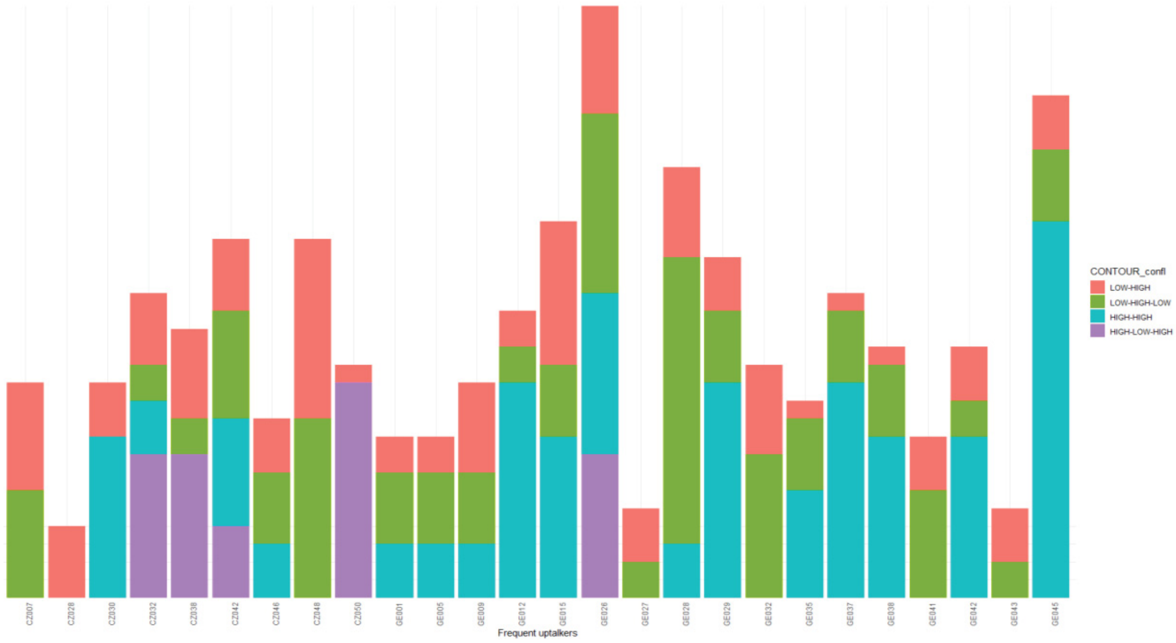


Figure 41. Frequent uptalkers with more than three uptalk tunes of the 560 word samples and their tune choices

Figure 41 also shows that all 26 frequent uptalkers use at least two to three different tunes throughout the 560 word samples. Speaker CZ028 uses only a LH-tune, while speakers CZ032, CZ042, and GE026 use all four tunes.

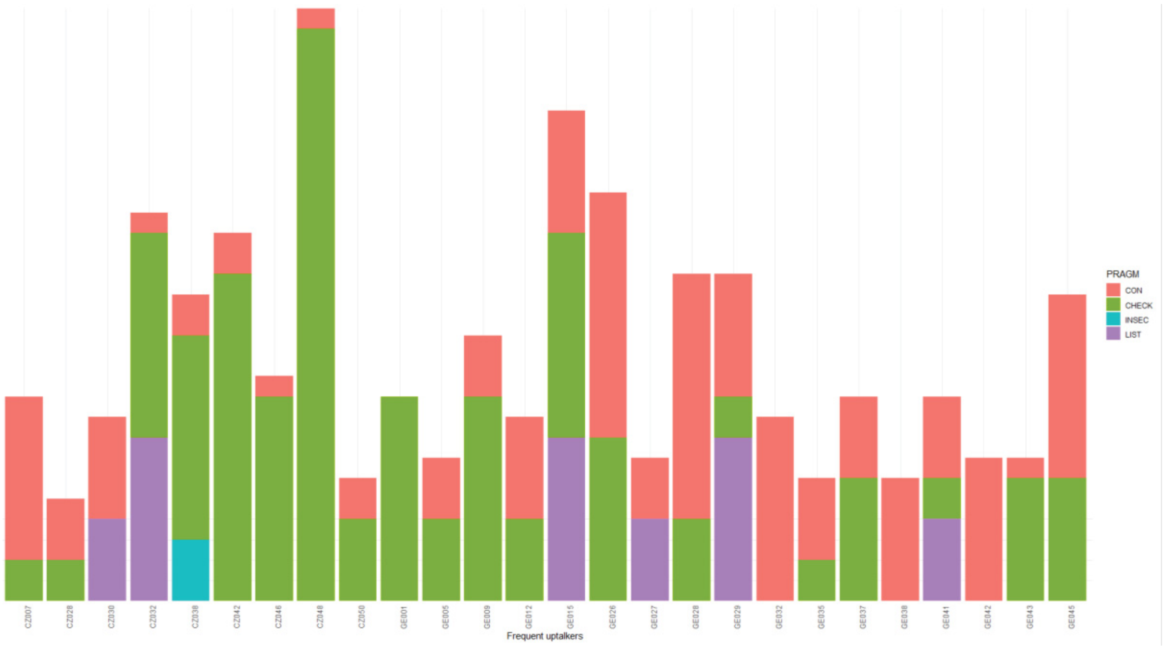


Figure 42. Frequent uptalkers with more than three uptalk tunes of the 560 word samples and the pragmatic functions used

Figure 42 demonstrates that most CzE speakers use the different uptalk tunes for checking, while the GerE speakers seem to exhibit a tendency to signal the wish to continue their utterances. Most speakers seem to alternate between two pragmatic functions (except GE001) and only very few speakers use three pragmatic functions (n=4). Speaker CZ038 is the only speaker who uses uptalk to signal insecurity.

Before the uptalk instances are analyzed phonetically, the context has to be considered more closely as a first step, i.e. the interviewer sex and variety of English, and the verbal behavior of the interviewer (backchanneling). High boundary tones have been found to invite backchanneling feedback by listeners (Allan 1990; Barry 2008; Gravano and Hirschberg 2009). Barry (2008: 208) found that uptalk tunes are most frequently followed by "acknowledging" turns from the interlocutor. Allan (1990: 124), calls these backchanneling devices "minimal responses" and found 33.33% of all uptalk instances to be followed by such responses (ibid.: 126). "When a speaker clearly has the floor for a longer turn at talk, a rise can function to hold it, often eliciting a backchannel cue that functions to ratify the speaker's right to continue (cf. Schegloff 1981)" (McLemore 1990: 4). Ogden and Routarinne (2005: 168) found that the most common responses to HRTs are discourse markers such as *mm* and *aha* and silent pauses. What type of response (full response, i.e. question/statement, or minimal response, i.e. backchanneling or silence) follows the uptalking instances is important, because it would support the theory of uptalk being used for floor holding rather than floor yielding and checking for comprehension, especially in semantically complex utterances (Guy and Vonwiller 1989: 25; Fletcher et al. 2005: 392). Therefore, the present analysis serves to verify whether the learners' uptalk instances are interpreted by the interviewers as checking, questions, or continuation (floor holding) devices and whether learners also get to hold the floor or whether the interviewer takes over. A lack of video recordings restricted the analysis to the lexical and prosodic domains of the speakers. Therefore, gestures and eye movement cannot be analyzed in the context of the present study. Table 40 provides an overview of uptalk instances by position in turn, backchanneling instances, pauses, and interviewer turns.

Table 40. Uptalk frequency by position and interviewer interaction

Factor		**Women (%) n=292**	**Men (%) n=29**
Position in turn	IP-medial (n=143) IP-final (n=178)	n=130 (44.52%) n=162 (55.48%)	n=13 (44.83%) n=16 (55.17%)
Backchanneling	No (audible) response Backchanneling	n=154 (52.74%) n=138 (47.26%)	n=14 (48.28%) n=15 (51.72%)
Interviewer intervention/speaker turn - before uptalk	No previous interruption Statement/question precedes uptalk instance	n=219 (75%) n=73 (25%)	n=19 (65.52%) n=10 (34.48%)
Pause type and duration by interviewee - after uptalk	Breathing or other Silent pause No pause	n=146 (50%) n=82 (28%) n=64 (21.92%)	n=16 (55.17%) n=8 (27.59%) n=5 (17.24%)
Interviewer intervention/speaker turn - after uptalk	No interruption/comment Interruption/followed by question/comment	n=283 (96.92%) n=9 (3.08%)	n=29 (100%) n=0

The same distribution of position in turn, backchanneling and interviewer turn before and after the uptalk instance and which type and duration of pause follows after the uptalk instance are very similar in the female and male speakers. It is almost equally divided between final and non-final uptalk positions within an IP. For the most part, the uptalk instances are not preceded or followed by any interviewer questions/statements. In addition, whether instances of uptalk were followed by audible backchanneling or not is almost 50/50. The ggplots in Figure 81 (see Appendix) show that uptalk instances are followed by audible interviewer backchanneling 0-100% of the time and 50% on average (median). The ggplots in Figure 82 (see Appendix) show that non-uptalking ips are followed by audible interviewer backchanneling 0-63% of the times and 22% on average (median). Comparatively, uptalk instances seem to be accompanied by backchanneling more often than non-uptalking instances. However, this depends on the regional variety and the interviewer sex. For the GerE and MurE female-female and CzE male-female conversations audible backchanneling after uptalk instances occur about 25% of the time. The uptalking Czech females in female-female conversations receive no backchanneling most of the time (median=0). However, when talking to the male non-native (Czech L1) interviewer, both Czech males and females receive 75% audible backchanneling feedback after uptalk instances. The ggplots (Figures 81 and 82 in the Appendix) show that the non-native male CzE speaker backchannels a lot more in general, also in non-uptalking ips. In the GerE corpus, however, the female-male conversations elicit more backchanneling after uptalk instances than in female-female conversations and the difference in non-uptalking instances is not as large as between the female-male and female-female conversations.

It seems that whether backchanneling is used or not does not only depend on the tonal patterns used, but more likely on the preferences and sex of the interviewer. The use of audible backchanneling might encourage the learners to use more uptalk in order to obtain verbal feedback. Other interviewers might make use of other backchanneling devices such as nodding, which can, however, not be tested in the present study. This depends on the personality of the interviewer and/or their understanding of the task. For instances, the female BrE interviewer from the GerE corpus generally kept backchanneling to a minimum in order to optimize the quality of the recording and/or not to disturb the speaker (information obtained from personal communication with the interviewer). Other interviewers, on the other hand, intervened a lot, such as the Belgian English speaker in the LOCNEC recordings.

In addition to the backchanneling behavior, the interviewees' pausing behavior after the uptalk instances was also investigated. 50% of the time the interviewees use breathing pauses after the uptalk instances, which might indicate that they intend to continue speaking. This claim will be further analyzed in the following. Ford and Thompson (1996: 157) found that a change in turn appears in AmE when intonation, syntactic, and pragmatic completion coincide. In SBrE, Wichmann and Caspers (2001) found that a high level tune (H* %) is perceived as a strong cue to floor-holding, regardless of syntactic completion, and falls are interpreted by listeners as floor-yielding.

Excluding the GerE and CzE groups, very few of the small number of uptalk instances in the other corpora are followed by interviewer questions or comments. However, in the CzE and the GerE samples one can see that interviewers only intervene if the pause duration is longer (i.e. one second longer than the usual pauses) (see Figure 83 in the Appendix). The pause durations after non-interrupted utterances is between 0.10-0.79 seconds long. The CzE and GerE speakers are extremely similar with 0.41 seconds on average (median) for the CzE and 0.45 for the GerE speakers. SpE and SAmE speakers have slightly longer pause durations, i.e. MadE (0.55 seconds), MurE (0.59 seconds), and SAmE (0.52 seconds). Those uptalk instances that were followed by interviewer interventions were 1.41 seconds long in the CzE utterances, and 1.33 seconds long in the GerE, MadE (1.16 seconds), and NBrE utterances (0.18 seconds). Thus, the longer the pause is, the more likely it is that the interviewer intervenes. An extreme high tone at the end of an IU accompanied by a shorter pause does not invite interviewer comments or questions. Figure 84 (see Appendix) shows that breathing pauses are generally longer in most regional varieties than silent pauses, and other types of pauses such as "ant", "intr", etc. have the longest pause durations.

72.27% (n=232) of the uptalk instances occurred during interviews conducted by a female interviewer and 27.73% (n=89) were conducted by a male interviewer. In total there are 13 different interviewers involved in the 83 interviews, two of which were male and the rest female. In the CzE corpus, half of the interviews are conducted by a female BrE interviewer, while the other half by a male CzE speaker. The female BrE speaker elicited uptalk by 11 speakers (26 uptalk instances from 11 interviews), and the CzE interviewer elicited 71 uptalk instances from 15 interviews. In the GerE corpus, there were five interviewers: AmE_1 with 12 "uptalkers" (67 uptalk instances), BrE_1 with 11 (49 uptalk instances), AmE_2 with five (32 uptalk instances), BrE_2 with three (four uptalk instances), and AusE, the only male interviewer, with three (18 uptalk instances). The female interviewers backchanneled 40.09% of the time (n=93) within the 232 uptalk instances of backchanneling, while the male speakers backchanneled audibly 60 times (67.42%). Proportionally, the female interviewers give fewer minimal responses after uptalk instances than male speakers.

The following uptalk tunes were identified and ordered according to their frequency in the analyzed corpus samples (see Table 41):

Table 41. Distribution of uptalk tunes by REGION and SEX

		IP-final				IP-medial			ip-medial	
Variety	**SEX**	**L* L-H%**	**L* H-H%**	**H*H-H%**	**H* L-H%**	**L* H-**	**H*H-**	**H* L* H-**	**L* H***	**H* H-**
CzE	**F**	14	19	6	0	35	7	8	0	0
CzE	**M**	1	4	1	0	1	0	0	0	0
GerE	**F**	25	47	29	1	31	11	2	3	3
GerE	**M**	1	6	2	0	7	0	0	1	0
MadE	**F**	0	1	4	0	5	3	0	0	0
MadE	**M**	0	0	0	0	0	8	0	0	0
MurE	**F**	1	0	6	0	7	0	0	0	0
MurE	**M**	0	0	0	0	0	0	0	0	0
AmE	**F**	2	6	0	0	1	3	1	0	0
AmE	**M**	0	1	0	0	1	0	0	0	0
NBrE	**F**	0	0	0	0	0	0	0	0	0
NBrE	**M**	0	0	1	0	0	0	0	0	0
Total		44	84	49	1	88	32	11	4	3

As one can see from the overview in Table 41, IP-final and IP-medial uptalk tunes are similarly frequent (178 IP-final vs. 131 IP-medial). Only seven of the uptalk instances occurred ip-medially. Since there are only so few of the ip-medial uptalk instances, they will be conflated with the IP-medial tunes for the statistical analysis. Late onset uptalk tunes are more common in IP-medial position (L* H-) than in IP-final position (L* L-H%) and they seem to be most commonly used by GerE and CzE female speaker. In order to further analyze the distribution of uptalk tunes by regional varieties of English, inferential statistics are needed. For the inferential analysis, the following research question and statistical hypotheses will be addressed: RQ8: *Does the distribution of uptalk differ depending on the region, speaking style, age, topic, sex, interviewer sex, ip length, and the pragmatic function used?* The statistical hypotheses are the following:

H0 the distribution of the DV UPTALK (yes or no) does not differ depending on the levels of the predictors (IVs and their interactions): REGION, the SPEAKING_STYLE, AGE, TOPIC, SEX, INT_SEX, ip length, and PRAGM: R^2 =0.

H1 the distribution of the DV UPTALK (yes or no) differs depending on the levels of the predictors (IVs and their interactions): REGION, the SPEAKING_STYLE, AGE, TOPIC, SEX, INT_SEX, ip length, and PRAGM: R^2>0.

There were no directional hypotheses for RQ8 due to of the lack of previous research on uptalking tunes in non-native speech. To statistically test whether speakers opt for uptalk or not, and whether uptalking behavior correlates with any other linguistic or extralinguistic variables, a mixed effects logistic regression model (theoretical conditional $R^{2=}0.91$) was fit.

UPTALK (uptalk or non-uptalk) was used as the dependent variable (see Table 42 for results). REGION, PRAGM, SEX, AGE, TOPIC, SPEAKING_STYLE, and IM_LENGTH_S_MED were entered as fixed effects. Since Bell and Johnson (1997: 12-13) found women to uptalk more when the conversational partner is a woman, the INT_SEX was entered into the model as a predictor with an interaction with interviewee's sex. Since NSs do not use uptalk instances in the corpus samples as often as the learners do, only the learner data was tested (n=135 speakers).

Table 42. Binomial glmm for uptalk usage UPTALK ~ 1 + REGION + PRAGM + SEX * INT_SEX + AGE + TOPIC + SPEAKING_STYLE + IM_LENGTH_S_MED + (1 | SPKR)

Fixed effects:	**Estimate**	**Std. Error**	**z-value**	**p-value**	
(Intercept)	-0.63824	2.35431	-0.271	0.78632	
REGION_GerE	1.37084	0.52313	2.62	0.00878	**
REGION_MadE	-1.44751	0.69498	-2.083	0.03727	*
REGION_MurE	-0.87187	0.73432	-1.187	0.2351	
PRAGM_checking	4.44976	0.15667	28.402	< 0.001	***
PRAGM_emphasis	-17.46708	4471.9207	-0.004	0.99688	
PRAGM_finality	-17.39983	1367.78909	-0.013	0.98985	
PRAGM_handing over turn	-17.81287	1051.71	-0.017	0.98649	
PRAGM_insecurity	-0.8131	0.60282	-1.349	0.17739	
PRAGM_listing	-0.40113	0.18404	-2.18	0.02929	*
SEX_male	-1.51133	0.50482	-2.994	0.00276	**
INT_SEX_male	0.45175	0.53945	0.837	0.40236	
AGE	-0.12852	0.08217	-1.564	0.11779	
TOPIC_entertainment	-0.16187	0.25635	-0.631	0.52774	
TOPIC_experience	0.25713	0.24465	1.051	0.29326	
TOPIC_self	0.18731	0.13369	1.401	0.1612	
SPEAKING_STYLE_dia	-0.27968	0.11921	-2.346	0.01897	*
IM_LENGTH_S_MED	-0.62627	0.51416	-1.218	0.22321	
SEX_male: INT_SEX_male	0.90706	1.29295	0.702	0.48297	

Number of obs.=28,932, SPEAKERS=135, Random intercepts: Speaker SD (1.65); Log likelihood =-2,531.5, theoretical conditional $R^{2=}$0.91, delta conditional R^2 =0.56.

CzE females in monologues, talking about a country with a female interviewer and using the continuative function, are -0.64 times less likely to use uptalk than a non-uptalking tune. SPEAKER accounts for 1.65 of the distribution. The GerE speakers are 1.37 times more likely to use uptalk in such a context than the CzE speakers (p=0.00878**). For the SpE speakers, it is less likely to use uptalk in such contexts but this observation is only statistically significant for the MadE speakers (β=-1.45, p=0.03727*). The results of the glmm also reveal that uptalk is most commonly used to signal the pragmatic function of checking (β=4.45, p<0.001***) and the least likely to be produced in lists (β=-0.4, p=0.02929*). While uptalk might appear (but less likely) in continuation, insecurity, and listing functions, it never appears in emphasis, finality, and handing over turn functions. Therefore, the estimates and standard errors are high for these pragmatic functions in the model. Male speakers are also less likely to use uptalk compared to female speakers (β=-1.51, p=0.00276**). For older speakers, the usage of uptalk becomes less likely, although the difference is not significant. The predictors TOPIC and intermediate-phrase length do not show any significant results, as well as the interaction between the interviewer and interviewee sex. The model does predict, however, that uptalk is less likely to be produced in dialogic speech (β=-0.28, p=0.01897*). The predictor REGION in the model is highly significant ($X^2(3)=28.76$, p<0.001***). Q-Q plots have, however, determined that the model is not distributed normally, due to the low occurrence of uptalk tunes compared to non-uptalking ones. Therefore, the results have to be interpreted with caution. Nevertheless, some general trends can be observed in uptalking and non-uptalking behavior. Adding fluency variables to the uptalk model, did not improve model fit and also did not show any significant results. All in all, H0 of RQ8 was rejected for the following variables: REGION, PRAGM, SEX, and SPEAKING_STYLE, while H1 was rejected for AGE, ip length, TOPIC, and INT_SEX.

A second mixed effects logistic regression model (theoretical conditional $R^2=0.90$, logLik=-2166.3, speaker intercept: 1.141, number of observations: 22,009, number of speakers: 103) including only the complete learner data was fit to answer RQ11g: *Does uptalk become less or more frequent depending on the length of stay abroad and the variety of English the learners have been exposed to?* The final model was the following: UPTALK ~ 1 + REGION * STAB + PRAGM + SEX * INT_SEX + AGE + TOPIC + SPEAKING_STYLE + IM_LENGTH_S_MED + YOEAS + YOEAU + NO_OFL + COUNTRY_confl + (1 | SPKR). If the learner variables and the interaction REGION*STAB are taken into account with the complete learner data from the CzE, GerE, and MadE speakers, the following significant results can be observed. GerE speakers produce significantly more uptalk instances compared to the CzE intercept (β=2.91, p=0.00623**). Uptalk is more frequent in the checking function (β=4.42, p<0.001***) and less frequent in lists (β=-0.46, p=0.02236*). Uptalk is also less frequent in speech produced by males ((β=-1.11, p=0.01336*) and more frequent when the interviewer is male (β=1.32, p=0.00837**). Also, older speakers are less likely to produce uptalk instances (β=-0.19, p=0.02260*). Interestingly, the learners who have been abroad to Ireland produce the most uptalking instances ((β=3.12, p=0.00986**). None of the other variables and their interactions turned out to be significant. In sum, while the length of a stay abroad does not have an effect on the productions of uptalk instances, only the country the learners have been to does. However, as with the model above, the confidence intervals for this model are too high and the findings have to be interpreted with caution.

4.4.1 F0 range measures of uptalk rises

Basic descriptive statistics for the different f0 measures of the uptalk tunes are listed in Tables 43-45 in absolute Hz values for f0 level, f0 min, and f0 max, STs for f0 span, and seconds for duration, broken down by REGION (CzE, GerE, and SpE). Only speakers who produce at least three uptalk instances were added to descriptive and the inferential analyses, which leaves only uptalk instances produced by female CzE and GerE speakers (and two MurE speakers for the descriptive results). There are, however, two GerE male speakers who produce uptalk frequently but they were also left out of the analysis (except in Figure 85 in the Appendix for a visual comparison with the female speakers).

Table 43. Descriptive statistics for f0 range measures of uptalk tunes: CzE corpus (female speakers only, n=9)

	N	Min	Mean	Max	SD	Median	IQR
Tune low (Hz)	70	132	183	227	21.82	182	37
Tune/rise peak (Hz)	70	207	272	422	39.44	264	46
Tune level (Hz)	70	181	228	316	27.99	222	40
Tune span (ST)	70	4.16	9.83	20	4.30	8.79	7.3
Tune duration (s)	70	0.53	1.23	2.22	0.41	1.12	0.51
Rise start (Hz) onset	70	127	195	248	25.26	198	37
Rise span (ST)	70	1.95	5.75	12.08	2.31	5.82	3.2
Rise duration (s)	70	0.22	0.45	0.92	0.16	0.41	0.19

Table 44. Descriptive statistics for f0 range measures of uptalk tunes: German corpus (female speakers only, n=17)

	N	Min	Mean	Max	SD	Median	IQR
Tune low (Hz)	137	107	187	281	23.69	188	26
Tune/rise peak (Hz)	137	212	302	446	50.15	300	65
Tune level (Hz)	137	159	245	363	30.55	244	42
Tune span (ST)	137	0.00	9.79	23.37	5.05	9.28	5.6
Tune duration (s)	137	0.50	1.92	102	8.61	1.09	0.52
Rise start (Hz) onset	137	148	199	302	26.02	196	29
Rise span (ST)	137	1.32	7.15	15.61	2.93	6.87	3.6
Rise duration (s)	137	0.20	0.47	1.23	0.17	0.43	0.18

Table 45. Descriptive statistics for f0 range measures of uptalk tunes: Murcia corpus (female speakers only, n=2)

	N	Min	Mean	Max	SD	Median	IQR
Tune low (Hz)	13	165	196	230	21.83	201	37
Tune/rise peak (Hz)	13	237	309	385	46.14	314	66
Tune level (Hz)	13	203	253	296	30.22	259	52
Tune span (ST)	13	3.24	9.95	24.51	4.92	8.77	2.4
Tune duration (s)	13	0.47	0.85	2.11	0.43	0.82	0.33
Rise start (Hz) onset	13	175	211	278	28.81	214	35
Rise span (ST)	13	1.96	6.53	10.74	2.62	7.60	3.7
Rise duration (s)	13	0.13	0.31	0.46	0.10	0.28	0.16

The f0 level of uptalk tunes of the learners is usually between 107 and 446 Hz. Figure 85 (see Appendix) shows the typical uptalk tunes of the frequent uptalkers broken down by REGION and SEX (only two GerE male speakers). Figure 85 (see Appendix) shows that many GerE speakers have very extreme peaks. The male speakers (only GerE and CzE) show that the CzE speakers overall have a higher f0 level than the other male speakers, but no huge differences in uptalk tune can be observed. One can see, however, that the male uptalk tunes are shallower.

As far as rise span in STs is concerned, similar results in line with previous research on uptalk can be found, i.e. men generally having narrower f0 spans. Considering all learner data (not the trimmed data), rise span in ST of CzE women is 5.96 vs. 5.38 for men, and 7.06 for GerE women and 6.28 for men, and MadE women have a span of 6.65 and 7.10 for MurE women. For comparison, Tyler (2019) found a mean rise span of 3.97 for men and 4.04 for women for AmE speakers in uptalk tunes. Surprisingly, the AmE women in the present study have a mean rise span of 3.97 STs (SD=2.74), exactly the same value as the AmE men in Tyler's (2019) study. However, the SD can vary by two to three STs and would thus be within the range of the learners, who seem to have comparatively high mean rise span values.

Comparatively, the uptalk tunes are fewer than non-uptalking instances in learner speech. However, the extreme rises might make them more noticeable to listeners and can, therefore, possibly be rated as less native-like (cf. Götz 2013). Therefore, the learners' uptalk tunes, produced by the frequent uptalkers only, are analyzed in more detail, in order to uncover realizational differences between the two learners groups. A comparison with NSs is not possible, because not enough uptalk instances were found in the data set of the present study. In the theory it was hypothesized that the learners of English will exhibit different phonetic realizations of uptalk tunes depending on the variety of English they have been exposed to. Furthermore, a by-sex analysis is not feasible because of the low number of male uptalkers in the data set. Therefore, the following analyses are based on nine CzE and 17 GerE female speakers. The statistical analysis tested whether POSITION, AGE, PRAGM, SPEECHACT, TIME_IN_RECORDING, INT_SEX, and TUNE_PATTERN are significant predictors of f0 level and span measures of uptalk tunes. The durational measures are analyzed in Section 4.4.2. Different mixed effects models were run with the following dependent variables:

1) F0 at onset (F0 minimum) in Hz
2) F0 at tune low (F0 minimum) in Hz
3) F0 at peak (F0 maximum) in Hz
4) F0 level (mean) in Hz of tune
5) F0 span of rise (difference between F0 min and F0 max) in STs

The f0 level of the rise and the f0 span of the entire tune were not included as dependent variables in the analysis because the results of the mixed effects model with other variables were almost identical. For all models the fixed effects were: REGION (CzE, GerE, MurE), POSITION (non-final, final), AGE, PRAGM (checking, continuation, listing, insecurity), TUNE_PATTERN (LH, LHH, HH, LHL), SPEAKING_STYLE (dia, mono), SPEECHACT (answer, statement), INT_SEX (female, male), TIME_in_RECORDING (beginning, later), and TTR (Guiraud-index). SPEAKER was included as a random intercept. The intercept for all the following models is equal, i.e. female CzE speakers with a LH-tune in a continuative function in non-final position in statements in dialogues when the interviewer is female and the uptalk instances appear in early parts of the recordings. When necessary, the models were trimmed to achieve a normal distribution of the residuals of the respective models. The following research question was tested: RQ9: *Does the f0 level and span of uptalk differ depending on the region, speaking style, age, topic, sex, interviewer sex, position in the IP, speech act, TTR, and pragmatic function used?*

H0 the f0 level/span of the uptalk rises does not differ depending on the levels of the predictors (IVs and their interactions): REGION, SPEAKING_STYLE, PRAGM, AGE, PRAGM, TUNE_PATTERN, POSITION, SPEECH ACT, INT_SEX, and Guiraud-index: R^2 =0.

H1 the f0 level/span of the uptalk rises differs depending on the levels of the predictors (IVs and their interactions): REGION, SPEAKING_STYLE, PRAGM, AGE, PRAGM, TUNE_PATTERN, POSITION, SPEECH ACT, INT_SEX, and Guiraud-index: R^2>0.

Since only the CzE and GerE data is considered in this analysis, RQ11h (*Do uptalk f0 measures change after a stay abroad and the variety of English the learners have been exposed to?*), will be answered at the same time and the learner variables (AGE, YOEAS, YOEAU, STAB, NO_OFL, and COUNTRY) will be considered in the models. The final models for each f0 measure of the uptalk tunes only include the main effects.

The first (trimmed to exclude residuals deviating by more than 2 SDs, final model: F0_MIN_ONSET ~ TUNE_PATTERN_confl + POSITION (1 | SPKR)) model with f0 min at rise onset (F0_MIN_ONSET; β=193 Hz for intercept) determined that the following factors are significant (see Table 46 and Figure 43 for main effects): HH-tunes are higher in their f0 minimum at rise onset (β=20.35 Hz, p<0.001***) compared to low-high tunes. Low-high-high tunes are also higher in their f0 at rise onset (β=15.35 Hz, p<0.001***). F0_MIN_ONSET is lower in final position (β=-10.80 Hz, p<0.001***). Speaker accounts for 14.16 of the variance and there is a residual variance of 14.57 left. The results of model comparisons with anova indicate that there was a main effect of TUNE_PATTERN_confl (X^2(47.85)=3, p<0.001***), and POSITION (X^2(14.14)=1, p<0.001***). None of the other factors reached significance.

Table 46. Results for mixed effects model with F0 min at rise onset in Hz. Final model: F0_MIN_ONSET ~ TUNE_PATTERN_confl + POSITION + (1 | SPKR)

Fixed effects:	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	193.045	3.437	56.168	<0.001	***
TUNE_PATTERN_conflLHH	15.349	3.043	5.045	<0.001	***
TUNE_PATTERN_conflHH	20.346	3.146	6.468	<0.001	***
TUNE_PATTERN_conflHLH	-10.796	5.773	-1.87	0.063055	.
POSITION_final	-11.061	2.889	-3.829	<0.001	***

Number of obs.=197, SPEAKERS=26, Random intercepts: Speaker SD (14.16), Residual SD (14.57); Log likelihood =-833.7, conditional $R^{2=}$0.57.

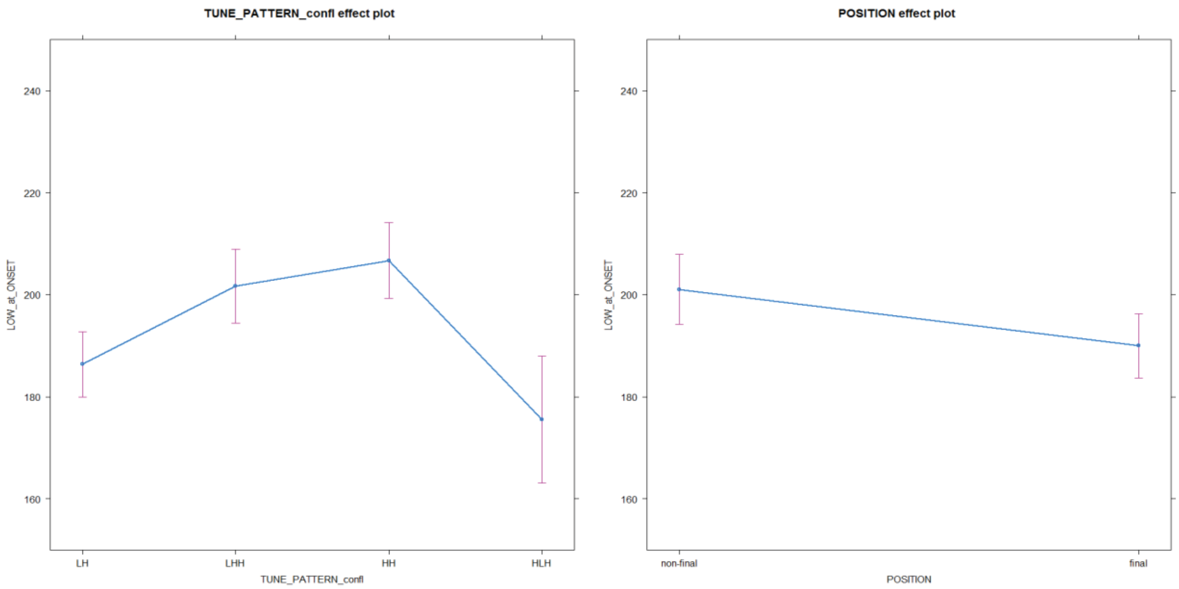


Figure 43. Effect plots for mixed effects model results of f0 minimum at rise onset in an uptalk tune

In a second step I modeled f0 height at tune low (β=168 Hz) as a function of variety of English (REGION) and the other predictors from the first model (see Table 47 and Figure 44 for main effects). The model (trimmed to exclude residuals deviating by more than 2 SDs, final model: F0_MIN ~ TUNE_PATTERN_confl + SPEAKING_STYLE + SPEECHACT + (1 | SPKR)) predicts that HH-tunes (β=16.49, p<0.001***) are higher in their f0 min than LH-tunes, which was to be expected considering the ToBI labels of H vs. L. In monologues f0 min at tune low is by 13 Hz higher (p<0.001***) and answers (SPEECHACT) are by 11 Hz higher (p=0.004088). If the interviewer is male, f0 min at tune low is 36.27 Hz higher (p=0.008586**). Speaker accounts for 12.16 of the variance (residual variance=11.99). The results of model comparisons with anova indicate that there was a main effect of TUNE_PATTERN_confl (X^2(40.93)=3, p<0.001***), SPEAKING_STYLE (X^2(12.24)=1, p<0.001***), and SPEECHACT (X^2(8.27)=1, p=0.004024**). None of the other factors reached significance.

Table 47. Results for mixed effects model with F0 min at tune low in Hz. Final model: F0_MIN ~ TUNE_PATTERN_confl + SPEAKING_STYLE + SPEECHACT + (1 | SPKR)

Fixed effects:	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	168.8947	4.5999	36.717	<0.001	***
TUNE_PATTERN_conflLHH	0.7995	2.1583	0.37	0.711507	
TUNE_PATTERN_conflHH	16.4921	2.5738	6.408	<0.001	***
TUNE_PATTERN_conflHLH	-1.1797	4.697	-0.251	0.801962	
SPEAKING_STYLE_mono	13.3177	3.7471	3.554	<0.001	***
SPEECHACT_answer	11.3787	3.9143	2.907	0.004088	**

Number of obs.=197, SPEAKERS=26, Random intercepts: Speaker SD (12.16), Residual SD (11.99); Log likelihood =-796.3, conditional $R^{2=}$0.59.

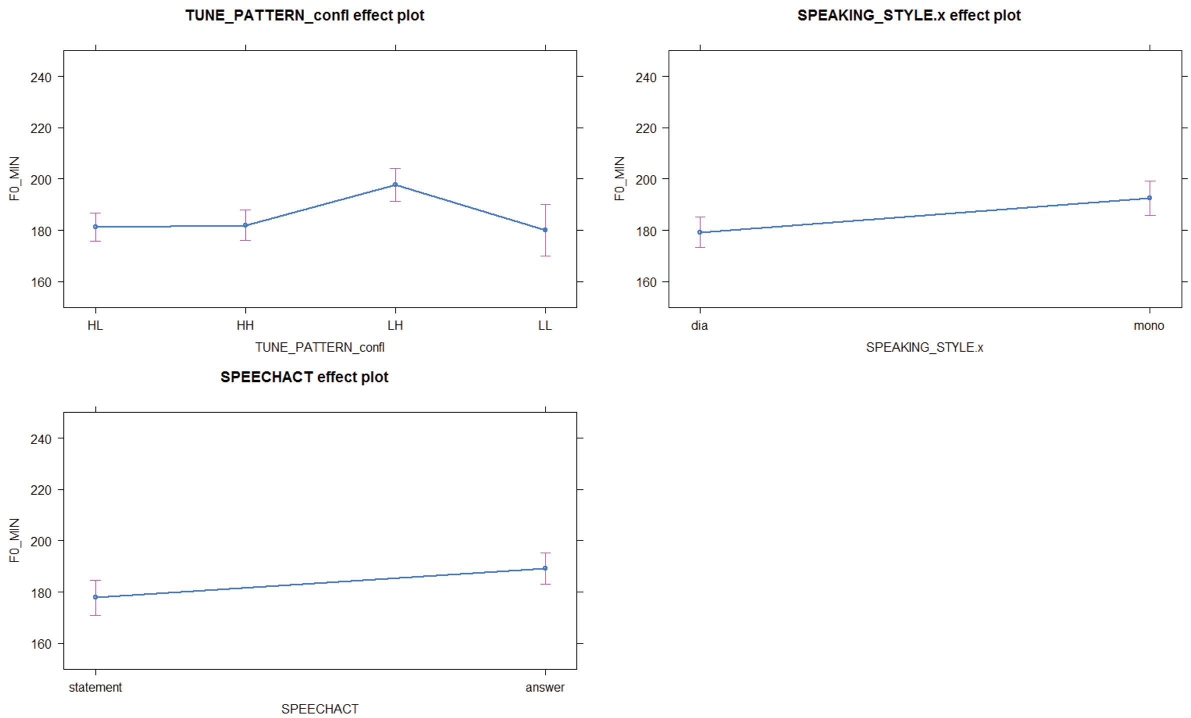


Figure 44. Effect plots for mixed effects model results of f0 minimum at tune low in an uptalk tune

For f0 of the rise peak none of the predictor variables reached significance. Another mixed effects model (trimmed to exclude residuals deviating by more than 2 SDs, final model: F0_MEAN ~ TUNE_PATTERN_confl + (1 | SPKR)) with f0 level of the entire uptalk tune (β=235 Hz) as the dependent variable was fit (number of observations: 198, 26 speakers, conditional R^2=0.60, logLik=-877.2). The results show that HH-tunes have a significantly higher f0 level (β=13.46, <0.001***). The random intercept of SPEAKER accounts for 20.29 of the variance and there is a residual variance of 17.39. Thus, the only significant predictor for this particular model was TUNE_PATTERN_confl (X^2(24.05)=3, <0.001***). None of the other variables reach significance.

In a last step, rise span in STs (β=6.94 STs) was introduced to the model (trimmed to exclude residuals deviating by more than 2 SDs) (see Table 48 and Figure 45). The final model (F0_SPAN_RISE_ST ~ TUNE_PATTERN_confl + REGION * STAB + POSITION + (1 | SPKR)) results indicate that low-high-high (β=-2.4 STs, p<0.001***) and HH-tunes (β=-1.78 STs, p<0.001***) are predicted to have a narrower rise spans compared to LH-tunes (Plot 1 in Figure 45). In final position rise spans are wider (β=1.43 STs, p<0.001***) than in non-final position (Plot 2). The GerE with no stay abroad produce significantly wider rise f0 spans (β=4.77 STs, p=0.009379**) than those who have been abroad (Plot 3). Speaker accounts for 0.74 of the variance (residual variance=2.06). The significance of rise span predictors are the following: TUNE_PATTERN_confl (X^2(32.39)=3, p<0.001***), REGION*STAB (X^2(16.04)=5, p=0.006736**), and POSITION (X^2(11.88)=1, p<0.001***). None of the other factors reached significance.

Table 48. Results for mixed effects model with rise span in STs. Final model: F0_SPAN_RISE_ST ~ TUNE_PATTERN_confl + REGION * STAB + POSITION + (1 | SPKR)

Fixed effects:	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	6.9434	1.0748	6.46	0.001	***
TUNE_PATTERN_conflb_LOW-HIGH-HIGH	-2.3964	0.4322	-5.545	0.001	***
TUNE_PATTERN_conflHIGH-HIGH	-1.7769	0.4404	-4.035	0.001	***
TUNE_PATTERN_conflHIGH-LOW-HIGH	0.729	0.7605	0.959	0.338952	
REGION_GerE	-0.8676	1.3125	-0.661	0.514923	
STAB_NO_STAB	-2.4992	1.362	-1.835	0.077293	.
STAB_SHORT	-1.4209	1.1293	-1.258	0.221802	
POSITION_final	1.4254	0.4062	3.509	0.001	***
REGION_GerE:STAB_NO_STAB	4.7745	1.7149	2.784	0.009379	**
REGION_GerE:STAB_SHORT	2.4585	1.4093	1.745	0.094287	.

Number of obs.=197, SPEAKERS=26, Random intercepts: Speaker SD (0.74), Residual SD (2.06); Log likelihood =--430.2, conditional $R^{2=}$0.35.

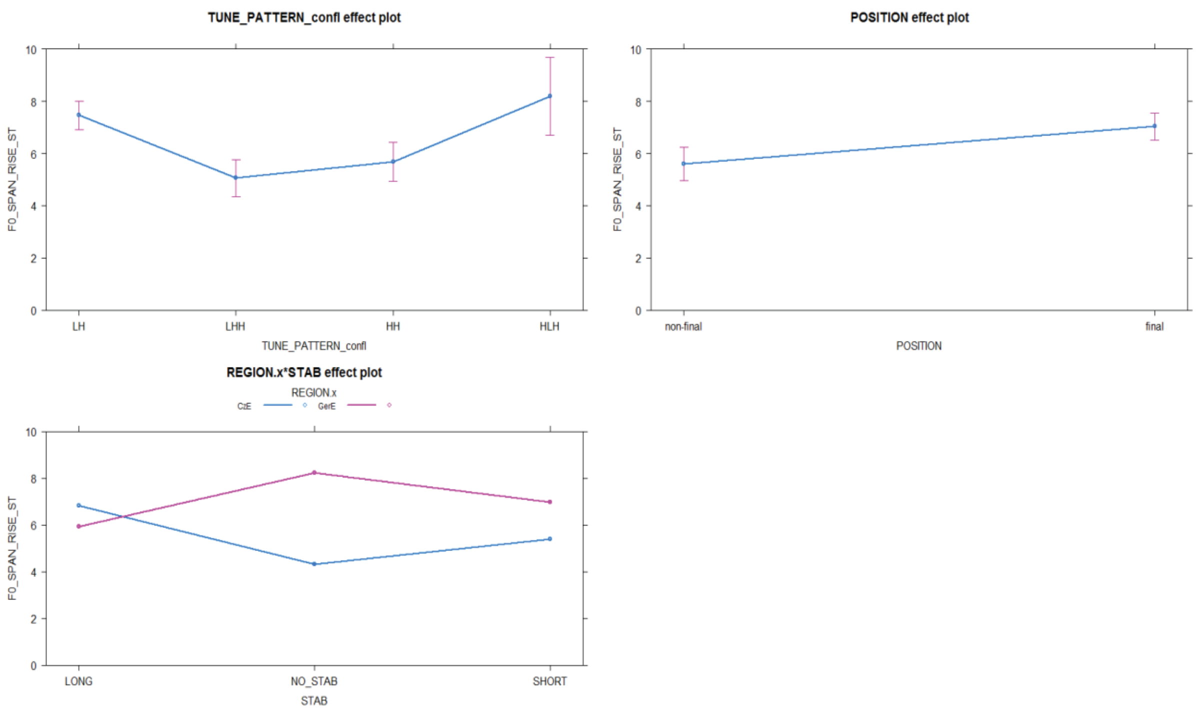


Figure 45. Effect plots for mixed effects model results of f0 rise span in STs in an uptalk tune

In sum, the most important predictors to determine f0 range differences in uptalk tune seem to be the tune pattern and the position of the tune. Even though the analyses return significant results, it has to be noted that there are only three speakers with long stays abroad (CZ038, GE009, and GE042) and five with no stay abroad (CZ028, CZ046, GE012, GE041, GE043), and all others experienced a short stay abroad (CZ007, CZ030, CZ042, CZ048,

CZ050, GE001, GE005, GE015, GE026, GE027, GE028, GE029, GE032, GE035, GE037, GE038, GE045). Even though the number is uneven, the analysis is not based on one instance per speaker but three or more uptalk tunes per speaker. Generally, the analysis revealed that there are no significant phonetic differences of uptalk tunes between the CzE and GerE speakers. Only for rise span, GerE speakers with no stay abroad are predicted to produce wider spans than CzE speakers. The position of the uptalk tune only plays a role to the related rise measures of f0 min at rise onset and f0 rise span, where f0 is lower at rise onset and f0 rise span is wider in final position. Speaking style and speech act only play a role for f0 min at tune low, where f0 is higher in monologues and answers. Interestingly, the pragmatic functions never reached significance in any of the models. No matter for what pragmatic function the uptalk tunes are used, no significant differences can be observed (marginally significant for checking function for f0 min at tune low). Indicating that the learners do not realize their f0 range in uptalk tunes to signal different pragmatic functions. None of the other factors reached significance. Therefore, H0 for RQ9 can only be partially rejected for some of the predictors.

4.4.2 Slope and duration of uptalk rises

The slope of the rise was measured as the f0 span of the rise divided by the duration of the rise in seconds. The f0 span of the rise was measured as the difference between the f0 height at the rise peak and the f0 low at the onset of the rise. The f0 span was converted to STs with the following formula: 12*log2(F0_MAX/F0_MIN_TUNE_LOW). The slope of the rise was then obtained by dividing the f0 span in STs (f0 at peak - f0 at onset) by the rise duration in seconds (onset to peak). The slope of the rise measure provides an indication of how fast the f0 changes during the production of the rise. In addition to the absolute slope of the rise, the relative slope was measured for every uptalk instance in relation to the slope of rise of the entire IP. Sometimes, however, there are multiple uptalk instances within one IP (reduplication). In these cases the uptalk instances were divided by the same IP slope. The relative slope of rise expresses the absolute slope of rise as a proportion of the total slope of an IP, which is more comparable across speakers and IPs (Di Gioacchino and Crook Jessop 2011; Levon 2018: 13). Thus, the slope of rise of each uptalk instance was divided by the overall slope of the encompassing IP in order to normalize the measure according to its prosodic environment which a speaker normally produces in the same context. Barry (2008: 170), for instance, found that SoCal and London speakers do not make use of their full f0 ranges in HRTs and they use significantly larger f0 ranges over the turn than they do for an HRT. Therefore, both absolute and relative slope of rise are analyzed in the present section.

The slope of the final rise (in Hz and ST) and the length (in seconds) of the final rise were analyzed for each uptalk tune. The following descriptive statistics are based on the untrimmed uptalk data (including also those speakers with less than 3 uptalk instances). Uptalk tunes were classified into three shape types, depending on their phonetic realization, i.e. a rise (n=246), sustained (n=74), and rise-plateau slump (n=1). In the present study, the learners mainly used a rising shape for their uptalk instances, which in my definition can include both a late or sudden steep rise, and an earlier steep rise. Flaig and Zerbian (2016) also found German learners of English to mainly prefer a continuous steep rise, as opposed to a sudden steep rise and a continuous flat rise. In the present data set, on average the length of these uptalk tunes ranges from 0.13 to 1.32 seconds with a mean of 0.45 seconds (median=0.41 seconds). The slope of the final rise usually averages to 231.18 HZ (16.77 STs) (median=187.02 Hz, 14.20 STs), with a range of 3.41 Hz (0.39 STs) to 869.61 Hz (67.02 STs). These extremely high values indicate a rapid change in f0, where an extreme peak value of up to 438 Hz is reached from a low at onset of 229 Hz within 0.25 seconds. The descriptive statistics are broken down by REGION and SEX in Table 66 (in the Appendix). For the female CzE speakers, the mean relative slope in STs is higher by 1.91 STs than the slope of rise in isolation, which means that if the context is considered in which the uptalk tunes occurs, the CzE speakers raise their pitch in uptalking tunes compared to their general f0 slope. In the GerE female speakers their slope of rise in STs is by 4.03 STs lower if the context is considered. For MadE females the mean is higher, and for MurE it is lower. Table 67 (in the Appendix) shows the results for the frequent uptalkers only (n=26 speakers); the differences to Table 66 (in the Appendix) still persist.

Correlations for rise duration and rise slope were run to test whether the two dependent variables were measuring different aspects of the rises. The effect plot in Figure 46 shows that there is a negative correlation between rise duration and rise slope, i.e. the longer the rise, the shallower the slope (Pearson's r=-.55).

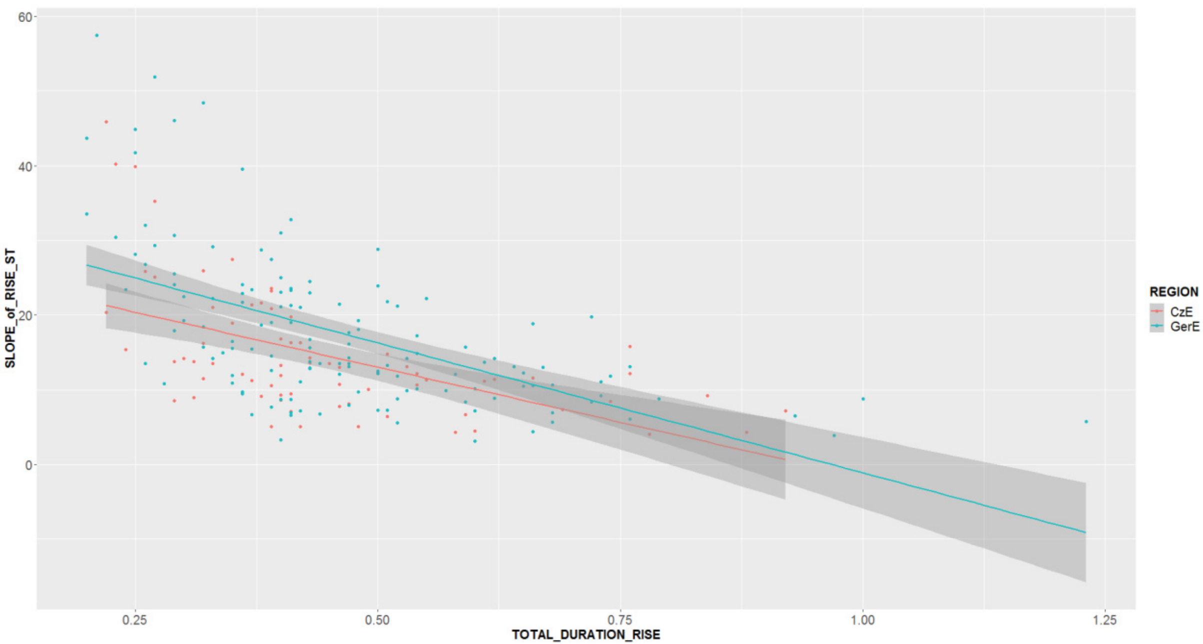


Figure 46. Effect plot for the correlation of rise duration and slope in STs by REGION

Visualizations (Figure 47 and 48) showed that CzE speakers produce steeper relative slopes rises (in STs and logged) which consequently means that their rise durations in seconds are lower than those of the GerE speakers. However, there is also considerable overlap between the two speaker groups for these two variables.

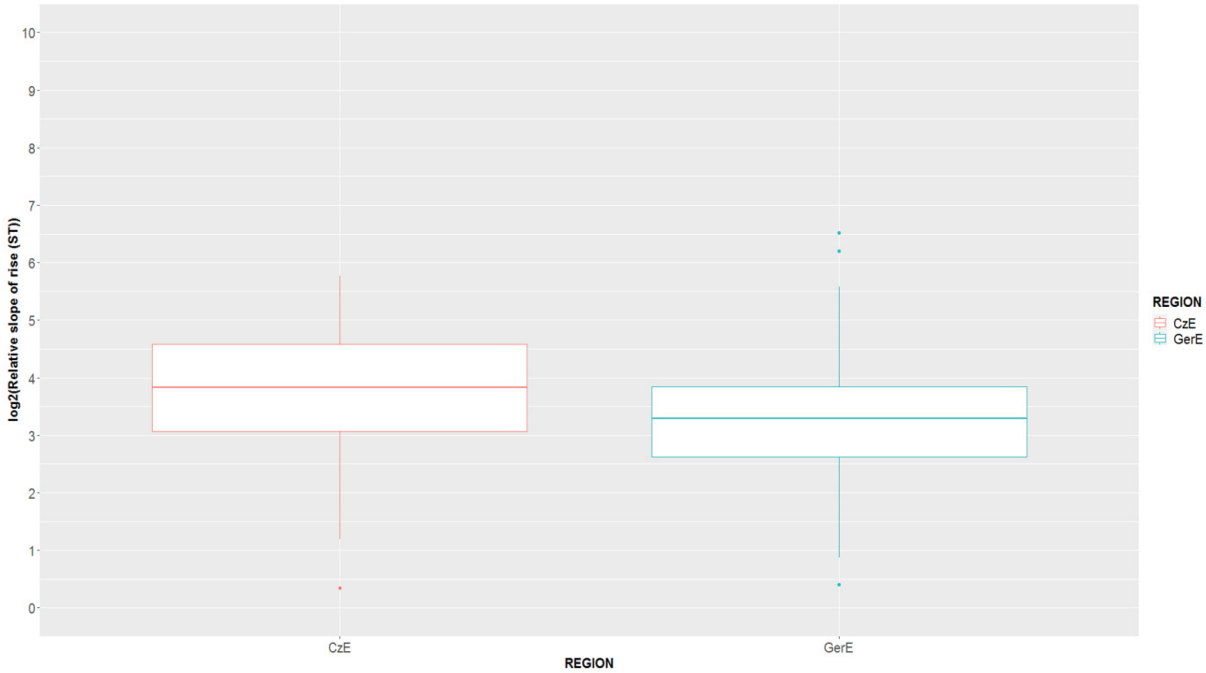


Figure 47. Boxplot for relative slope of rise in STs and logged values of CzE and GerE speakers

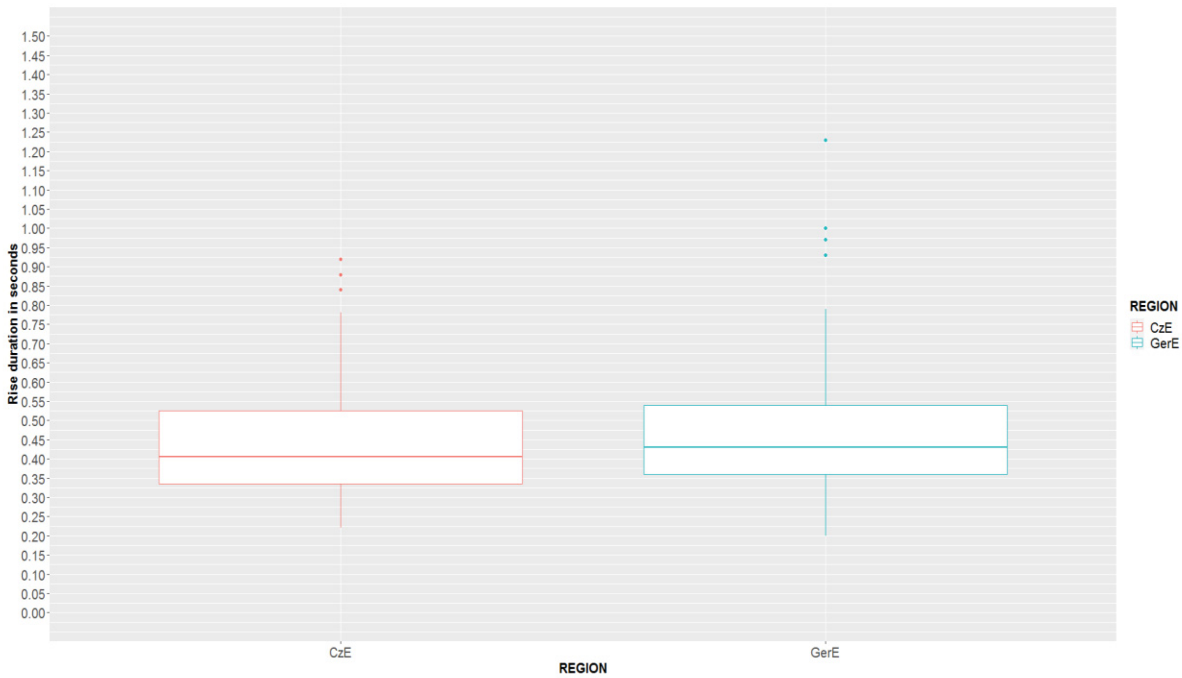


Figure 48. Boxplot for rise duration in seconds of CzE and GerE speakers

For the mixed effects models, the following statistical research questions and hypotheses were formulated: RQ10: *Are there differences in the relative slope of the rise in uptalk tunes in the different varieties of English? Does this differ depending on the tune shape, speaking style, age, interviewer sex, position in the IP, ip frequency, and pragmatic function?*

H0 the relative slope of the rise of the uptalk rises does not differ depending on the levels of the predictors (IVs and their interactions): REGION, SPEAKING_STYLE, AGE, PRAGM, TUNE_PATTERN, POSITION, ips, INT_SEX, and RISE_DURATION: $R^2 = 0$.

H1 the relative slope of the rise of the uptalk rises differs depending on the levels of the predictors (IVs and their interactions): REGION, SPEAKING_STYLE, AGE, PRAGM, TUNE_PATTERN, POSITION, ips, INT_SEX, and RISE_DURATION: $R^2 > 0$.

Since only the complete data of the frequent uptalkers of the CzE and GerE corpora were included, RQ11h (*Do uptalk f0 measures change after a stay abroad and the variety of English the learners have been exposed to?*) will be answered at the same time and all learner variables will be included in the mixed effects model. SPEAKER was included as a random intercept. Since the data deviated from normality, as verified by residual plots, the DV was logged (log2). Since previous studies found that the size of the rise is also correlated with their pragmatic function, i.e. with rises that signal floor-holding being produced with an f0 range twice as large as those for non-floor-holding statements (Ritchart and Arvaniti 2014), the pragmatic function was added to the mixed effects model. Previous research has shown that in AmE, BrE, and Canadian English, for instance, uptalk is not normally used to show hesitation or uncertainty (e.g. Shokeir 2008; Ritchart and Arvaniti 2014; Levon 2018). Warren and Daly (2005) found significant differences in rise slope in New Zealand English for different speaking styles, i.e. steeper slopes on narratives than on simple statements, as well as sex differences with women having steeper slopes than men. Therefore, the variable SPEAKING_STYLE was added to the model as well but was removed from the final model since it returned insignificant results and did not improve model fit. SEX could not be added as a predictor because the trimmed data set only includes female speakers. The relative slope of rise in STs was taken as the dependent variable.

The final mixed effects model (see Table 49, conditional $R^{2=}0.31$) demonstrates that LH-tunes are on average produced with 5 STs (logged scale) and only LHH-tunes are produced with a significantly shallower slope (β=-0.38 STs, p=0.015603*) (Plot 1 in Figure 49). Uptalk tunes produced to signal insecurity (PRAGM) are significantly shallower than those in the continuative function (β=-3.33 STs, p<0.001***) (Plot 2). The analysis also revealed that the slope of the rise and rise duration are highly negatively correlated, i.e. the longer the duration of the rise, the shallower the slope becomes (β=-2.13 STs, p<0.001***) (Plot 3). Armstrong et al. (2016) found the same result regarding these two variables. YOEAS (β=0.18 STs, p=0.026485*) lead to higher slopes (Plot 4) and YOEAU to shallower ones (β=-0.23 STs, p<0.001***) (Plot 5). There seems to be a tendency for learners with long no- or short stays abroad to produce shallower slopes, especially for the CzE speakers this difference is notable (all p<0.001***) (Plot 6). None of the other predictors reached significance and were therefore removed from the final model. Thus, the only significant predictors in this model, are PRAGM ($X^2(12.57)=3$, p=0.005674**), the duration of the rise ($X^2(27.16)=1$, p<0.001***), YOEAS ($X^2(4.94)=1$, p=0.0263*), YOEAU ($X^2(13.37)=1$, p<0.001***), and REGION*STAB ($X^2(26.28)=6$, p<0.001***).

Table 49. Results for mixed effects model with relative slope of rise in STs. Final model: log2(RELATIVE_SLOPE_UPTALK) ~ REGION * STAB + TUNE_PATTERN_confl + PRAGM + TOTAL_DURATION_RISE + YOEAS + YOEAU + (1|SPKR)

Fixed effects:	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	5.05742	0.93208	5.426	<0.001	***
REGION_GerE	-1.55956	0.46349	-3.365	<0.001	***
STABNO_STAB	-2.14472	0.45564	-4.707	<0.001	***
STAB_SHORT	-1.41667	0.36958	-3.833	<0.001	***
TUNE_PATTERN_confl_LHH	-0.38283	0.157	-2.438	0.015603	*
TUNE_PATTERN_confl_HH	-0.09235	0.16595	-0.556	0.578479	
TUNE_PATTERN_confl_HLH	-0.34278	0.31148	-1.1	0.272402	
PRAGM_checking	-0.15861	0.14038	-1.13	0.25984	
PRAGM_insecurity	-3.32802	0.95245	-3.494	<0.001	***
PRAGM_listing	-0.21606	0.31131	-0.694	0.488447	
TOTAL_DURATION_RISE	-2.12563	0.39451	-5.388	1.94E-07	***
Years.of.English.at.school	0.18001	0.08054	2.235	0.026485	*
Years.of.English.at.university	-0.22753	0.05928	-3.838	<0.001	***
REGION_GerE:STAB_NO_STAB	1.88042	0.56319	3.339	<0.001	***
REGION_GerE:STAB_SHORT	1.78529	0.45272	3.943	<0.001	***

Number of obs.=206, SPEAKERS=26, Random intercepts: Speaker SD (0.00), Residual SD (0.88); Log likelihood =-266.1, conditional $R^{2=}$0.3

In sum, H0 for RQ10 can only be rejected for the predictors PRAGM, YOEAS, YOEAU, and the interaction REGION*STAB, considering all other independent variables do not explain the variation in the data. The mixed effects model explains only 31% of the variance in the data, and overall GerE speakers tend to produce a shallower relative slope of uptalk tunes compared to the CzE speakers (β=-1.56 STs, p<0.001***).

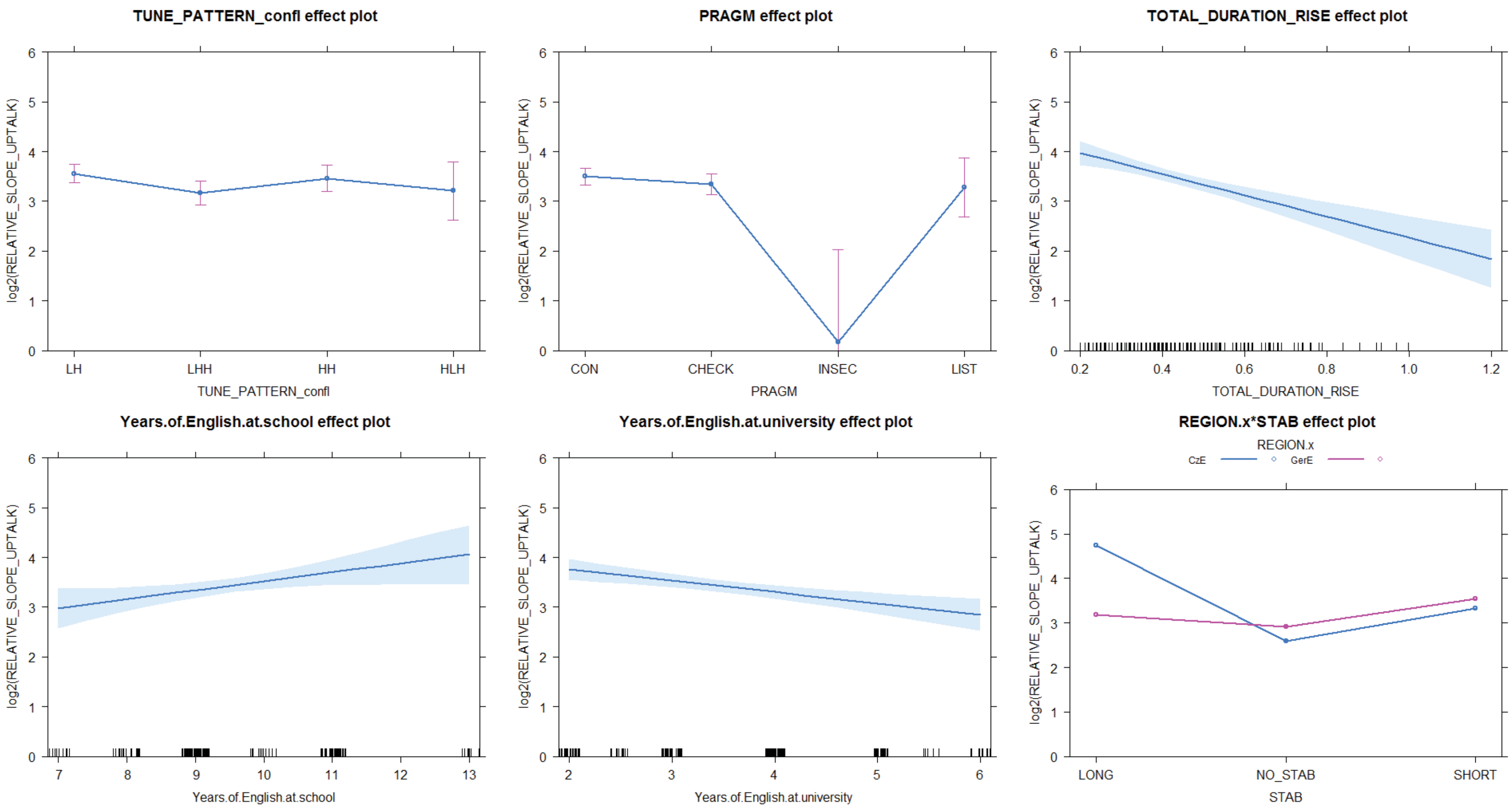


Figure 49. Effect plots for mixed effects model results of relative slope of uptalk tunes in STs

4.5 Summary

Section 4 has analyzed the four main prosodic features (tones and tunes, intonational phrasing, f0 range, and uptalk) and revealed that the advanced learners significantly differ from the NSs across nearly all prosodic features. The findings are summarized in Section 5.1, however a short preview is given here, sorted according to native-like, non-native-like, and L1-based native-like and non-native-like differences:

Native-like prosodic features all learners have in common:
1. Tone frequency
2. F0 level of HL-tunes
3. All learners have positive DID-scores and therefore entrain to their interlocutors

Non-native-like prosodic features all learners have in common:
1. High edge tones and tunes - more frequent
2. Speech rate per IUs - slower speech rate
3. ip length in seconds - longer ips
4. F0 span of HL-tunes - narrower f0 span
5. F0 span for HH- and LH-tunes - wider f0 span
6. F0 level for HH- and LH-tunes - higher f0 level

Native-like L1-based results:
1. SpE speakers produce native-like IU frequencies

Non-native-like L1-based differences:
1. IU frequency (CzE and GerE females with diverging results)
2. Most high edge tones by GerE and CzE speakers
3. SpE speakers produce narrowest f0 span for HL-tunes and widest for HH- and LH-tunes
4. SpE speakers do not entrain as much as other learners
5. MurE most divergent on f0 range - produce widest and highest f0 on LL-tunes
6. MurE produce more high pitch accents
7. MadE produce slowest speech rate within IPs
8. CzE produce slowest speech rate within ips

Each prosodic feature (including the various measures) was analyzed from a descriptive as well as inferential perspective (as far as the available data allowed). After the presentation of the descriptive results (mainly via ggplots), each dependent variable was entered into a statistical model. Each statistical model was chosen on the basis of the nature of the dependent variable, i.e. logistic regressions for numeric values with only one data point per speaker, mixed effects models for multiple numeric values per speaker, and binomial logistic regressions for categorical data with two levels. The results were presented with the help of tables and effect plots. The NBrE speakers were taken as the intercept for most models where a comparison between NSs and NNSs was made. This type of comparison was conducted first. In a second step, the CzE speakers were taken as the intercept whenever only the NNSs were compared to each other. Section 4.4 was the only section that did not include an analysis of NSs vs. NNSs productions due to data sparcity for the NS data.

5. Discussion of the results

As stated in the introduction, the main purpose of this study was to examine whether the distribution, frequency, and function of L2 prosodic features differ between the non-native and native speakers, and whether there are differences according to L1 background. In addition, other (extra-)linguistic factors besides L1 influence that can explain the prosodic variation were also investigated. I hypothesized that the learners from different L1 backgrounds will exhibit differences specific to their L1 and, thus, exhibit features of L1 transfer or -influence in their L2. Thus, the productions of the learner groups were hypothesized to be distinguishable in their prosody from each other (NNS vs. NNS), even though some similarities/universals might exist compared to the native controls (NNS vs. NS). Returning to one of the questions posed at the beginning of this study, i.e. whether learners and NSs establish prosody differently, it is now possible to state that the three learner groups of English differ significantly from NE speakers in nearly all aspects of prosody investigated in this study. However, there are differences in the degree of deviances based on L1 and speaker sex. Due to the large number of hypotheses of the present study, all hypotheses are listed in Table 50 and summarized, interpreted, and a global prosodic profile of the learner groups is discussed. The structure of the discussion is the same as in previous sections, i.e. a by-feature discussion is given. Even though the prosodic features are discussed in separate sections, references to results from other sections are made frequently, i.e. several prosodic features may be discussed together. Then, the results are put into perspective via reference to the LILt framework and other SLA theories, such as CAT and the negotiation of meaning. In a final step, an L2 prosodic model based on the results is proposed and implications for language teaching and speech technology are discussed.

5.1 Global discussion of the prosodic features

Table 50 shows each hypothesis as stated in Section 3.7.4 and details whether these hypotheses were confirmed, partially confirmed, or rejected, by indicating the intonational dimension addressed and the R^2 value or C-score of the respective statistical model that addressed the individual hypotheses. This is followed by a summary of the prosodic profiles of the three learner groups and three native groups.

Table 50. Summary of all hypotheses and results (D=distributional, F=functional, and R=realizational)

Prosodic features	Hypotheses	Intonational Dimension	Confirmed	Partially Confirmed	Rejected	R^2	Expla-nation
Tones and Tunes	RQ1 - H1: NNSs produce more tones	D			**X** (no differ-ences found)	.78	Adjusted R^2 in linear regression
	RQ11a:[76] *More or fewer tones after stay abroad?*	D			**X** (no differ-ences found)	.75	Adjusted R^2 in linear regression
	RQ2 - H2: NNSs produce more high edge tones and high-register tunes	D	**X** (all learners but especially CzE and GerE)			.76	C-score in multiple logistic regression
	RQ11b: *Fewer high edge tones after stay abroad?*	D	**X** (only true for CzE)		**X** (only true for GerE and MadE)	.71	C-score in multiple logistic regression
Intonational Phrasing	RQ3 - H3: NNSs produce more IUs	D		**X** (only for GerE females for IP level)	**X** (for most learners no differ-ences found)	IPs: .52 ips: .66	Adjusted R^2 in linear regression
	RQ11c: *Fewer IU breaks after stay abroad?*	D			**X**	IPs: .59 ips: .71	Adjusted R^2 in linear regression
	RQ4 - H4: NNSs produce slower speech rate within IUs	R	**X**			IPs: .37 ips: .23	Conditional R^2 in mixed effects model
	RQ11d: *Faster speech rate per IU after stay abroad?*	R			**X**	IPs: .17 ips: .092	Conditional R^2 in mixed effects model

[76] Some of the research questions include questions because no hypotheses were made beforehand due to a lack of previous research. All other hypotheses include statements because they were formulated as directional hypotheses in Section 3.7.4. The RQs in Table 49 are abbreviated versions of the RQs from Section 3.7.4.

Prosodic features	**Hypotheses**	**Intonational Dimension**	**Confirmed**	**Partially Confirmed**	**Rejected**	**R^2**	**Expla-nation**
	RQ5 – H5: NNSs produce native-like f0 level	R	**X** (for HL- tunes only)		**X** (for LH- and HH- tunes)	.86	Conditional R^2 in mixed effects model
	RQ5 - H6: F0 level position-sensitive	R	**X**			.86	Conditional R^2 in mixed effects model
	RQ5 - H7: F0 level differs by ip length	R	**X**			.86	Conditional R^2 in mixed effects model
	RQ5 - H8: F0 level higher in monologues	R			**X** (result n.s.)	.86	Conditional R^2 in mixed effects model
F0 range F0 level	RQ5 - H9: F0 level higher in checking function	R & F	**X**			.86	Conditional R^2 in mixed effects model
	RQ5 - H10: F0 level higher in post-interactions	R	**X**			.86	Conditional R^2 in mixed effects model
	RQ5 - H11: F0 level higher the more topics are initiated	R			**X**	.86	Conditional R^2 in mixed effects model
	RQ5 - H12: F0 level higher the more hesitations are used	R			**X** (result n.s.)	.86	Conditional R^2 in mixed effects model
	RQ11e: *Lower f0 level for LH- and HH- tunes after stay abroad?*	R			**X**	.84	Conditional R^2 in mixed effects model

Prosodic features	Hypotheses	Intonational Dimension	Confirmed	Partially Confirmed	Rejected	R^2	Expla-nation
	RQ6 - H13: NNSs produce narrower f0 span	R	**X** (for HL-tunes only)		**X** (for LH- and HH-tunes)	.38	Conditional R^2 in mixed effects model
	RQ6 - H14: F0 span position-sensitive	R	**X**			.38	Conditional R^2 in mixed effects model
	RQ6 - H15: F0 span wider in longer ips	R	**X**			.38	Conditional R^2 in mixed effects model
	RQ6 - H16: F0 span wider in dialogues	R	**X**			.38	Conditional R^2 in mixed effects model
	RQ6 - H17: f0 span wider in checking function	R & F	**X**			.38	Conditional R^2 in mixed effects model
	RQ6 - H18: F0 span wider in post-interactions	R	**X**			.38	Conditional R^2 in mixed effects model
F0 range F0 span	RQ6 - H19: f0 wider the more topics are initiated	R			**X**	.38	Conditional R^2 in mixed effects model
	RQ6 - H20: f0 span wider the more hesitations are produced	R			**X**	.38	Conditional R^2 in mixed effects model
	RQ11f: *Wider f0 for HL-tunes & narrower for LH- & HH-tunes after STAB?*	R			**X**	.23	Conditional R^2 in mixed effects model
	RQ7 - H21: NNSs & NSs both entrain to respective interlocutors	R	**X**			.24	Conditional R^2 in mixed effects model
	RQ7 - H22: NNSs will entrain more	R			**X**	.24	Conditional R^2 in mixed effects model
	RQ7 - H23: Females entrain more	R		**X** (true but depends on gender-pairings)		.24	Conditional R^2 in mixed effects model

Prosodic features	Hypotheses	Intonational Dimension	Confirmed	Partially Confirmed	Rejected	R^2	Expla-nation
	RQ8 - H24: Females produce uptalk more often	D	**X**			.91	theoretical conditional R^2 in binomial glmm
	RQ8 - H25: Uptalk more frequent in monologues	D	**X**			.91	theoretical conditional R^2 in binomial glmm
	RQ8 - H26: Uptalk more frequent in longer IUs	D			**X** (no effect found)	.91	theoretical conditional R^2 in binomial glmm
Uptalk	RQ11g: *Fewer or more uptalk after STAB and variety of English NNSs have been exposed to?*	D		**X** (NNSs who have been to Ireland produce more uptalk)		.90	theoretical conditional R^2 in binomial glmm
Uptalk	RQ9 - *F0 level and span differences of uptalk tunes?*	R		**X** (small phonetic differences but GerE tend to produce wider rise spans)		.35 - .59	Conditional R^2 in mixed effects model
	RQ10 - *Rise slope and duration differences of uptalk tunes?*	R	**X** (CzE produce steeper slopes and shorter rise durations)			.31	Conditional R^2 in mixed effects model
	RQ11h: *Change in uptalk f0 after STAB and variety of English NNSs have been exposed to?*	R	**X** (CzE speakers with long STAB produce steepest slopes & GerE produce widest rise spans with no STAB)		**X** (no effect of COUN-TRY variable)	F0: .35 - .59 Slo pe: .31	Conditional R^2 in mixed effects model

In sum, 19 out of 36[77] individual RQs and hypotheses were confirmed, 18 were rejected, and four were partially confirmed. To summarize Table 50, the similarities and differences between non-natives (NNS vs. NNS) will be considered, and a NS-NS contrast will also be made. To get back to the original four research questions:

1) What are the structural and functional features of prosody in the spoken interlanguage of advanced learners of English with different L1 backgrounds?
2) Can universal features be observed across different language families?
3) a) To which extent do these learners diverge from the native speakers' prosodic patterns or adopt language-appropriate values in spontaneous speech?
   b) If the learners deviate from native productions, what are possible reasons? What is the sociolinguistic distribution of these features?
4) What makes the prosodic features of foreign language learners so foreign-sounding?

On the basis of the present study, it can now be said that NNSs have the following prosodic features (structural and functional) in common as compared to the NBrE norm (RQs1-3a):

1. Tones and tunes: <u>Tone frequency is native-like for all NNSs</u>
2. Tones and tunes: Tonal categories: Higher frequency of high edge tones and high-register tunes and learners are less likely to use high pitch accents (except MurE speakers)
3. Intonational phrasing: Slower speech rate (WPS) per IUs
4. Intonational phrasing: Longer ips in seconds
5. F0 range: <u>F0 level of HL-tunes close to target norm</u> (HH- and LH-tunes are higher)
6. F0 range: F0 span of HL-tunes narrower compared to native f0 span (HH- and LH-tunes are wider)
7. F0 range: <u>All learners have positive DID-scores and therefore entrain to their interlocutors</u>

From the above features that all learners have in common, only <u>three</u> out of seven reached native-like values and the learners deviate from the NSs for all other features. Furthermore, prosodic features were found that differ depending on the NNS group (RQs1-3b):

1. Tones and tunes: GerE & CzE speakers produce the most high edge tones/tunes
2. Tones and tunes: MurE speakers produce more high pitch accents
3. Intonational phrasing: GerE females produce more IPs phw
4. Intonational phrasing: CzE females & MurE males produce fewer IPs phw
5. Intonational phrasing: CzE females produce fewer ips phw
6. Intonational phrasing: CzE speakers produce the longest ips in words
7. Intonational phrasing: MadE speakers produce the slowest speech rate within IPs
8. Intonational phrasing: CzE speakers produce the slowest speech rate within ips
9. Intonational phrasing: SpE speakers produce the shortest ips in words
10. Intonational phrasing: <u>SpE speakers produce native-like IU frequency</u>

[77] Most RQs had multiple results, i.e. for the different regional varieties and/or IUs (IPs vs. ips).

11. Uptalk: There seems to be small phonetic and phonological differences in uptalk tunes produced by CzE & GerE speakers: GerE speakers (especially with no stays abroad) tend to produce wider f0 spans in STs in uptalk tunes. There is a tendency for CzE speakers (especially with long stays abroad) to produce steeper relative slopes of rises in STs of uptalk tunes and, consequently, they produce slower rise durations in seconds as compared to GerE speakers. All other f0 measures of uptalk tunes did not exhibit any significant differences between CzE & GerE speakers
12. F0 range: SpE speakers produce the narrowest f0 span for HL-tunes and widest for HH- and LH-tunes
13. F0 range: Besides producing a higher f0 level for HH-, and LH-tunes, as the other learner groups, MurE speakers also produce the highest and widest f0 for LL-tunes
14. F0 range: SpE speakers tend to entrain their f0 the least to that of their interlocutors
15. Pragmatic functions: CzE & GerE speakers use prosody more often to signal a wider variety of pragmatic functions
16. Pragmatic functions: GerE speakers use extremely low f0 to signal a handing over of turn (MadE speakers also exhibit this trend) and an extremely high f0 to signal continuation in order to possibly compensate for smaller chunks on the IP level
17. Pragmatic functions: CzE speakers use high-register tunes the most in a continuative function to signal that they intend to hold the floor possibly due to their lower production of IU breaks

In sum, from the 17 individual differences listed above, the intonational phrasing level seems to be the most problematic with the different learner groups clustering at different ends of the extremes (compared to NS-norm). While the productions of the CzE speakers can be characterized by containing fewer IUs, the longest ips in words, and the slowest speech rate within ips, the GerE female speakers produce more IPs, while the SpE speakers do not exhibit any problems with IU frequency (except MurE males), they do produce the slowest speech rate within IPs (MadE). Also, for tonal categories the CzE/GerE and MurE speakers seem to favor opposite trends: while the former more often opt for high edge tones, the latter produce more high pitch accents. While the SpE speakers are the only ones who produce a native-like prosodic feature (besides tone frequency and f0 level for HL-tunes, which all learners have in common), i.e. IU frequency (RQ3a), they seem to exhibit additional deviances from native-like f0 range (see items 12-14 above), thus showing mostly difficulties at the realizational level. Overall, the CzE and GerE speakers seem to be the most similar on almost all distributional, realizational, and functional dimensions of prosody investigated in the present study. Differences in f0 range (level and span) were lower than expected, indicating that the CzE and GerE learners are at a quite advanced proficiency level with regard to this prosodic feature. Moreover, only small differences were found in the production of uptalk tunes (f0 range, slope, and duration) between CzE and GerE speakers. Additionally, to answer RQ3b, it is mostly CzE and GerE female speakers, who deviate the most from native speech, however, this point will be taken up again in Section 5.5.2.

Finally, RQ4, cannot be answered with absolute certainty, because a perception study would be needed to find a more conclusive answer. Nevertheless, it seems that the uptalk tunes are the prosodic features that create the biggest impression of foreign-soundingness in the samples analyzed in the present study (from my point of view). However, this is more likely due to a combination of all of the prosodic features analyzed. In fact, their interrelatedness will be discussed in more detail in the following sections.

The present study also uncovered similarities and differences in the prosody of NSs, which is important to consider if one wants to take the NS baseline as the target model. Therefore, the prosodic features with virtually no differences between NSs were the following:

1. Tones and tunes: Tone frequency is similar across the native varieties
2. IP frequency: No significant differences between NSs (except SAmE females)
3. ip length in seconds: All three native varieties produce around 1.5 seconds long ips
4. ip length in words: All three native varieties produce six words on average per ip
5. IU speech rate (WPS): No significant differences between NSs
6. Tune choice: H* L-L% tune most frequent for all NSs
7. F0 span: No significant differences between NSs for most tunes (except LL-tunes, which are narrower in both AmE varieties)
8. Uptalk: Virtually no uptalk usage
9. Pragmatic functions: No significant differences in the use of pragmatic functions with certain tunes, the HL-tune is mostly used to signal continuation

Prosodic features with significant differences between NSs were the following:

1. Tones and tunes: SAmE females use tones that are more typical of NBrE males
2. High pitch accents: SAmE and AmEO are less likely to use high tone as a pitch accent, although H* is the most common tone used as a pitch accent (something they have in common with NNSs)
3. High edge tones: SAmE and AmEO are more likely to use a high edge tone (something they have in common with NNSs)
4. IP frequency: SAmE female speakers are predicted to produce fewer IPs phw
5. ip frequency: SAmE female speakers are predicted to produce more ips phw and SAmE males are predicted to produce fewer ips phw
6. F0 level: SAmE speakers produce a higher f0 level for HL-tunes than NBrE speakers, SAmE LL-tunes are lower than NBrE LL-tunes
7. F0 span: SAmE and AmEO produce narrower f0 span in LL-tunes

In sum, the SAmE speakers seem to deviate considerably on almost all intonational dimensions from NBrE speakers, while AmEO speakers seem to have more in common with NBrE speakers yet, at the same time, also cluster with SAmE speakers. However, the deviances may be explained by differences in speaking style, with the SAmE speakers having more of a narrative monologic speaking style and AmEO and NBrE speakers being involved in mostly dialogic speech.

5.2 Discussion of tones and tunes results

The learners' tonal choices have been shown to differ from native speech. The preferred ip tune was found to be a high-low tune (H* L-) for all corpora, native and non-native. However, certain situations and speaker pairings lead to higher frequencies of HH- and LH-tunes in non-native speech, while the NSs mostly stick to HL-tunes. All learners were found to use more high edge tones and high-register tunes, with GerE and CzE producing most high edge tones and tunes. This finding will be discussed from a functional perspective in Section 5.2.2. However, the differences in tone and tune choice between NSs will be discussed first in the following section.

5.2.1 Tones and tunes in native speech

All NSs clearly stick out with a predominantly HL-tune choice for all pragmatic functions (mostly continuative function used), the SAmE speakers seem to be the most deviant of the NSs in that they pattern more with the NNSs, when it comes to the use of high edge tones. The NSs in the present study share many similarities in their prosodic productions, one of them being that high boundary tones (H%) are rather the exception. However, some differences between NSs were also detected, i.e. SAmE speakers are more likely to deviate from the most common tune pattern (H* L-L%) and to use L* pitch accents early in their ips instead of a high pitch accent. Since SAmE female speakers were also found to produce significantly more ips than NBrE speakers, it is only logical that they consequently produce more tones as well, considering a positive correlation was found between a higher frequency of ips and tones. Since the SAmE samples were characterized by a narrative speaking style, mainly in monologic form, the speakers broke up their utterances into smaller and more manageable chunks for their listeners and, therefore, exhibited the tendency to produce more tones at the same time (not statistically significant). In addition, the results revealed that a faster speech rate lead to a lower production of tones, while more silent pauses consequently lead to more tones. This is due to the internal structure of ips, which at least needs to contain one pitch accent and a phrase accent and an additional boundary tone if it is an IP-final ip. While speech rate per IU has not been found to yield significant differences between the three native varieties, the AmE groups tended to produce longer IPs in seconds. However, the AmE speakers (SAmE speakers in particular) exhibited the tendency to produce more filled and silent pauses, as well as elongations of syllables, which could explain the IP length differences and ultimately the tendency for SAmE to produce more tones phw (not statistically significant). This also could explain why high edge tones are more likely in AmE (SAmE particularly).

The result that SAmE speakers are more likely to produce high edge tones may also point at social as well as regional differences between the NS groups. For instance, previous research found that there also seems to be a dialectal difference in marking the connection to successive ips, in that SAmE female speakers frequently use high phrase accents (H-) compared to Midland females (Clopper and Smiljanic 2011: 243). However, female speakers of both dialect groups in Clopper and Smiljanic's (2011) study preferred the low-high edge tone (L-H%) combination, which in their study suggested a gender difference in marking the relationship to successive IPs to mark continuation. Thus, a typical intermediate utterance in read speech from a SAmE female speaker would be a low-high pitch accent with a high phrase accent (L*+H H-) and the typical intonation phrase would include a low-high pitch accent, a low phrase accent and a high boundary tone (L*+H L-H%). Clopper and Smiljanic (2011: 244) suggest that this preference may reflect female speaker's attitudes and emotion toward the content of the utterance or they could be interpreted as uncertainty. The present study also found AmE speakers to be more likely to use high edge tones than NBrE speakers, which is also likely to be conditioned by gender differences, i.e. female speakers using it more often than males to mark continuation.

Thus, while the NSs are extremely similar to each other in their tone frequency and tonal choices, the speaking style and regional differences have to be kept in mind, which can lead to different preferences in tone and tune choice. These differences are particularly important when NS samples are used as a benchmark for comparison to the productions of non-native speech. In comparison to the NBrE speakers, the NNSs all deviate on almost all prosodic levels investigated in the present study. However, when compared to the AmE (especially SAmE), more similarities can be found, i.e. longer IPs, more frequent or fewer IU breaks (differences by IU level and gender-based differences), and more low pitch accents instead of high ones, as well as more high-register tunes.

Additionally, recent theories on the meaning of intonation claim that differences in the meaning of tunes should also correspond to differences in their phonological form (Brazil et al. 1980; Ladd 1983, 2008; Gussenhoven 1984, 2004; Bolinger 1986; Pierrehumbert and Hirschberg 1990; as summarized in Levon 2018: 17). Some empirical research, however, contends that the difference between AmE and BrE, for instance, lies mainly in the phonetic realization of phonological forms, rather than using different phonological forms for distinct pragmatic functions (Fletcher et al. 2005: 401-402; Barry 2008: 273). On the other hand, no phonetic characteristics were found to be used with distinct pragmatic functions in London English (Levon 2018). In line with Levon's research, the present study did not find any phonetic or phonological differences in the use f0 for distinct pragmatic functions in the different native varieties of English. In general, however, the findings of the present study are consistent with the viewpoint of prosody-meaning relations, since a trend of different phonological forms corresponding to certain pragmatic functions was observed, i.e. ips signaling a handing over of turns are used with lower and narrower f0, while continuative utterances are used with higher and wider f0. The learners, however, showed additional variation within these form-meaning relations of prosody, as the next section explains.

5.2.2 Tones and tunes in non-native speech

Especially for the learner data there was considerable phonetic realization within these phonological categories, and many different phonological forms were used for a variety of pragmatic functions, i.e. the same phonological form was used in many different pragmatic functions. Arvaniti (2011: 8), summarized results of previous research and comes to the conclusion that "no useful generalizations about either intonational meaning or intonational form can be made on the basis of contour shape and its relationship to broad functional effects", since "the same melody can lend different pragmatic nuances to different utterances, while the same meaning can be expressed by superficially different-looking contours" (Arvaniti 2011: 6-7). She then concludes that, "Overall, then, melodies do not appear to have specific functions, and indeed attempts to describe the melodies of specific pragmatic nuances, such as irony, have proved unsuccessful (e.g. Bryant and Fox Tree 2005)" (Arvaniti 2011: 7). Nevertheless, as Section 2.5.3 (tones and tunes in native and non-native speech) has made clear, simply describing over- or underuse in non-native speech in comparison to native speech (Granger 2002) is no longer sufficient in contemporary learner corpus research.

Contextual and pragmatic features have to be taken into account as well (cf. Romero-Trillo 2018: 116). A more contemporary theory is prosodic pragmatics, which is a field that analyzes prosody and how it is used to express intentionality (Romero-Trillo 2012) and has been growing through the pioneering work by Romero-Trillo and his associates. Prosody has been described as a neglected area, especially from an L2 corpus pragmatic point of view (Fernández and Staples: 2021). In L2 pragmatic research, prosodic markers (PM) are often analyzed, i.e. discourse markers (e.g. *first*), general extenders (e.g. *and stuff*), and stance markers (e.g. *possibly*) (Fernández and Staples: 2021). Analyzing L2 speech from a prosodic pragmatic perspective helps learners to make their pragmatic intent more explicit. Failing to use prosodic patterns according to expectations can lead to negative perceptions of the speaker, i.e. signaling boredom or disinterest by using a monotonous speaking style (Kang 2010: 310) or it may lead to misunderstandings on the pragmatic level (e.g. Cenoz and García Lecumberri 1999: 4), or uncertainty, which has been the case with most explanations of L2 prosodic patterns. Previous research has found that NSs use rising tones for socially integrative purposes and as a face-saving function in situations that may potentially cause a conflict (Smith and Clark 1993: 36; Hewings 1995: 261; Ramírez Verdugo 2008; Pickering 2009; Pickering and Litzenberg 2011). For instance, instead of using a low tone, it may sound more polite and face-saving to say something to an interlocutor that he/she does not know with a rising tone. NNSs were not found to make use of rises associated with such functions in these studies and they were shown to opt for falls, which come across as more direct and less polite (Ramírez Verdugo 2008: 214). According to Hewings (1995: 262), NNSs do not have an awareness of the interrelatedness of intonation and social conventions. The learners in the present study, however, have demonstrated that they do have an awareness of the use of tones and social conventions, even if they 'overdo it' on the distributional and realizational level, as the next paragraph explains.

The two most common pragmatic functions (i.e. continuation (CON) and handing over of turns (HOT)) along with their tune patterns are summarized here in order to address RQ1 and RQs3a+b. For continuation both male and female NSs all show a clear distribution: The HL-tune is the most common tune, while for female speakers high-register tunes (HH and HF) are slightly more common and for male speakers the lower-register tunes (LL and LF) are slightly more common. The NNSs generally mirror this distribution but high-register tunes are produced more frequently to signal continuation, especially both female and male CzE speakers use the high-register tunes more often than a HL-tune. For HOT, a HL-tune is most frequent in native speech, where GerE and MadE speakers (both male and female) produce more low-register tunes. Especially the GerE males almost exclusively use LF-tunes with an extremely narrow f0 range to signal this function, while GerE females mostly use a wider f0 range but still with a low-register (LL-tune) (RQ3b). Thus, while all learners produced more high-register tunes in a continuative function, with CzE speakers being the most deviant, only GerE and MadE speakers produce more low-register tunes in the HOT function. Several linguists comment on the use of rises as cohesive devices and the tendency for them to appear in succession (e.g. Crystal 1969: 241; Britain and Newman 1992: 10; Bolinger 1989: 205; Wichmann 2000: 93; Linneman 2013). The learners may produce more high-register tunes in continuative utterances because they may feel the need to make it more clear that they intend to hold the floor and that their shorter IUs are to be seen as one coherent unit, while NSs may not feel the need to make coherence explicit via their tonal choice. Similarly, for the low-register tunes in HOT utterances, some learners seem feel the need to make it overly explicit that they intend to cede the turn. As discussed in the intonational phrasing section (see Section 5.3.2), the deviant usage of intonational phrasing phenomena is interconnected with their tonal-choice. Thus, while all NNSs speak slower (speech rate in WPS per IU), the GerE female speakers produce significantly more IP breaks and the CzE females produce significantly fewer IU breaks, and GerE and CzE female speakers produce significantly more high-register tunes. Thus, the deviant tune choice may be related to the GerE and CzE speakers' deviant intonational phrasing patterns, which go in different directions. The GerE female speakers produce smaller chunks of speech and need to make their pragmatic intent more explicit by overusing non-native-like and exaggerated patterns, either in an extremely low direction for HOT and an extremely high direction for CON utterances. The CzE female speakers on the other hand may use high-register tunes more because they intend to overly show that they intend to continue speaking. Thus, the CzE and GerE females use high-register tunes for similar strategies, i.e. to make their pragmatic intent overly explicit, while at the same time using different IU frequency patterns. This also goes hand in hand with the deviant f0 range (realizational level) of CzE and GerE learners (see Section 5.4.2), who produce higher and wider f0 for HH- and LH-tunes, which these learners may use in addition to the more frequent production of such tunes to make their pragmatic intent clear. Thus, in order to address RQ2, only GerE and CzE speakers seem to follow an extremely similar path in their prosodic productions (rather tune choice than IU frequency), while SpE speakers seem to exhibit more native-like patterns (on a distributional level at least). This suggestion is explained further in the next section, which discusses whether an L1 influence or developmental issues are more likely to explain the results.

5.2.3 L1 influence or developmental factors in tonal choices

Continuation rises were the most frequently used in the present study. High edge tones on declaratives in Czech, English, German, and Spanish may signal uncertainty or continuation (also called turn holding) (Selting 2010: 9; Wollermann 2012: 160). Based on previous research (e.g. Anufryk and Dogil 2009; Gut 2009; Anufryk 2012; Mennen et al. 2012), the preferred continuation tune was expected to be H* L-L% or H* L-H% in NE and L* H-H% or H* H-H% in GerE, thus showing L1 influence. The present study corroborated previous results (Chen 2003; Mennen et al. 2012: 2258), who found that German speakers preferred L* pitch accents, while BrE speaker preferred H* accents. Furthermore, in this study, it was also the non-native English speakers (mostly GerE) who overproduced L* while NBrE speakers used H* almost exclusively as a pitch accent. According to Mennen et al. (2012: 2258), English speakers prefer using the upper parts of their f0 range at the outset of an utterance more often, while German speakers tend to start their utterances rather low. That the CzE speakers use high boundary tones more frequently than the NBrE speakers could also be explained by L1 influence, because Chamonikolasová (2017: 66) found differences in non-terminal tunes in Czech and English with the reverse patterns of rises and falls, i.e. with Czech speakers using mostly rises and English speakers more falls. Previous research has found more native-like prosodic patterns in higher-aptitude and more proficient speakers, i.e. in the decrease of L*+H pitch accents and high edge tones (H-H%) with increasing proficiency, but also L1 transfer has been attested in L2 speech (e.g. Anufryk 2008, 2009, 2012; Anufryk et al. 2008; Anufryk and Dogil 2009; Mennen et al. 2010; Graham and Post 2018: 9-10). For the SpE speakers (MurE speakers), L1 influence could also be observed for the overproduction of high pitch accents, which also corroborates previous research (Ramírez Verdugo 2002: 122). However, SpE speakers were not found to produce more late peaks (L+H*) than any of the other native or non-native groups, and therefore no interference from the L1 could be attested. Regional variation is also a possible explanation for the distribution of tones and tunes. Previous research has shown that prosodic patterns transfer from the L1 regional accent to the L2 (e.g. Atterer and Ladd 2004). However, in the present study, regional varieties within the speakers cannot be accounted for in more detail, despite the fact that some speakers clearly speak with a regionally accented intonation in their L2. However, the differences may lie more in phonetic realizations such as prosodic alignment of syllables (e.g. Atterer and Ladd 2004) than in the tonal categories themselves. In sum, some L1 influence seems to be present in the case of the CzE, GerE, and MurE speakers.

At the outset of this study, it was hypothesized that more proficient learners may make use of the high edge tones because they have mastered the syntax and have a larger vocabulary to be able to afford to pay attention to their intonation. My assumption that more advanced learners will make use of rises for different pragmatic functions is mirrored in Kang et al. (2010). While no direct link with the speakers' TTR was found in the present study, the CzE and GerE speakers exhibited a similar frequency and shape of high edge tones and were additionally described as more advanced compared to the SpE speakers. A possible explanation of the higher frequency of high edge tones and non-native-like distribution and realization may be that rises generally develop more gradually than falling ones (e.g. Mennen et al. 2010), perhaps due to physiological difficulty (cf. Ohala and Ewan 1973) or because rises have various complex forms and meanings (cf. Crystal 1986; Grosser 1997; Snow and Balog 2002). The use of high edge tones as a possibly fossilized phenomenon will be discussed further in Section 5.5.3. Learners with fewer rises may be more focused on the language and not on their listeners (Pirt 1990 as cited from Kang et al. 2010: 562). Learners with lower proficiency levels do not use rises as a communicative device and may, therefore, be perceived as less proficient and less comprehensible (Kang et al. 2010: 562). While proficiency ratings for the Spanish learners do not exist, it becomes apparent from listening to the recordings and the prosodic results that they behave differently from the CzE and GerE speakers, which I suspect is mainly due to their lower proficiency level (my perception). Since the SpE speakers are the speakers with the narrowest f0 range (see Section 4.3), this shows that they are not prosodically as developed as the CzE and GerE speakers. Thus, while the SpE speakers very frequently pattern with the NSs in terms of their tune choice and intonational phrasing patterns, they definitely do not make use of f0 in a native-like manner. This may be seen as an avoidance strategy and by using less prosodically diverse speech, there are fewer opportunities for the SpE speakers to make deviant tonal choices. If they always speak in a monotonous way, then they will rarely deviate from the HL-tune. While the avoidance strategy may apply to the SpE learners when it comes to the HL-tune, it has to be noted, however, that they additionally produce the highest and widest f0 on HH- and LH-tunes (MurE females also produce highest and widest f0 on LL-tunes). However, on a distributional level SpE speakers use high edge tones less frequently than the other two learner groups but if they do use them, they produce them with the most extreme f0 in either direction (high or low). By producing the narrowest HL-tunes, SpE speakers may avoid varied f0 patterns because they may be focused on, for instance, lexical retrieval or other linguistic features (e.g. fluency or segmental features). For HH and LH-tunes the SpE speakers may avoid prosody that resembles their L1 (cf. Hulstijn and Marchena 1989) and may therefore produce more extreme f0 values. The MurE speakers were shown to produce the highest and widest f0 on LL-tunes (in addition to widest f0 span on LH- and HH-tunes), and since the Murcian native dialect has been described to be among the narrowest in f0 of all Spanish dialects (cf. Monroy and Hernández-Campoy 2015: 237), these f0 patterns can be seen as an attempt to deviate from f0 patterns that resemble their L1 dialect. Additionally, the fact that MurE female speakers are the only ones to entrain their f0 level less to that of their interviewers demonstrates that they are not perceptively and productively on the same level as the CzE and GerE speakers. The GerE and CzE speakers seem to be more advanced and seem to be on a 'prosodical exploration phase', where they 'overhit' the use of high and low tunes in certain pragmatic functions, but at least they are confident enough in their lexico-grammatical productions to venture out and to explore the use of high and low tunes to signal coherence (CON) and discourse functions (HOT). The pseudo-longitudinal study with learners who have been abroad and those who have not has shown that only the CzE speakers adopt more native-like tune patterns after a stay abroad, which is in contrast to previous research (Anufryk 2008; Anufryk et al. 2008; Anufryk 2009; Anufryk and Dogil 2009; Anufryk 2012: 291), who found that GerE female speakers make greater gains after a stay abroad. This may indicate that those learners who are restricted to a classroom setting for English language learning may need more time and practice (possibly through a semester abroad) to develop a more native-like use of high- and low-tunes. Why the GerE and MadE speakers seem to be 'resistant' to improving after a stay abroad needs to be investigated further in future research. A correlation analysis of the CEFR levels for the CzE speakers has additionally shown that with increasing proficiency, CzE speakers produce fewer high phrase accents and boundary tones (moderate correlation), while for the GerE speakers the correlations were extremely weak. These findings lend further support to my assumption that the more advanced learners are, the more they will venture out of their comfort zone, even if they do not reach native-like values, both on a distributional level (tone/tune choice) and realizational level (f0 range). It can be concluded, that while learners in the present study have acquired the compositional meaning of tones (i.e. that high edge tones are used to signal cohesion and continuation and low edge tones separation or completion from subsequent phrases), they use exaggerated versions, which may point towards a more realizational problem than a phonological one, which will be explained in the next paragraph.

While the present study treated the tones and tunes in the analysis as phonological features, a more detailed overview of the phonetic realizations of pitch accents, phrase accents, and boundary tones was provided in Sections 4.1.3-4.1.5. Previous research has found that while learners may show similar or equal phonological implementations of tones, there may be larger differences visible if the phonetic realizations of pitch accents are considered (e.g. Ramírez Verdugo 2006a: 25; Gut 2009: 248; Anufryk 2012). Indeed, the phonetic realizations of pitch accents yielded interesting results, with learners tending to use more varied pitch accent structures, which corroborated previous findings on SpE (cf. Ramírez Verdugo 2006a: 22). The CzE speakers produced the most different phonetic variations of H* accents compared to all groups, i.e. downstepped high pitch accents (!H*), upstepped high pitch accents (^H*), and regular high pitch accents (H*). While GerE and SpE realizational differences were already found in previous research, this is a novel finding for CzE. This finding could be seen as further evidence to support my assumption that learners try to adopt native-like prosodic patterns but are unsure about how to reach these f0 values and may try to vary them in an almost aimless manner until they reproduce an f0 contour that they think may resemble a native-like tune. Thus, learners may use f0 to try to vary their tune patterns in order to not sound monotonous and more native-like in their L2 productions. Nevertheless, the examples in Section 5.5 demonstrate that the f0 of high tones produced by the learners is indeed native-like but not produced in an aimless manner but is rather consistently ‘reproduced’ or ‘duplicated’ by the learners. This finding may also be seen as an L1 influence, considering certain pitch profiles that have been described for Slavic languages suggest that Slavic languages produce a more variable f0 as compared to Germanic languages (Andreeva et al. 2014a, b).

5.2.4 Summary

Learners of English have been shown to deviate in their tonal choice on a realizational, distributional, and functional level from native speech, which is also connected to their deviant intonational phrasing behavior as well as their f0 range productions. While L1 influence may be one possible factor that explains the use of reversed tune patterns compared to native speech, the proficiency level was suggested to be a more likely factor to determine tonal choice, due to the numerous universals found across the different L1 groups. It was suggested that the GerE and CzE speakers seem to follow a similar developmental path and find themselves in a ‘prosodic exploration phase’ with their use of high- and low-tunes being part of their developmental phase. The SpE speakers’ utterances are characterized by a seemingly more native-like tonal choice, which, however, may be more likely due to the speakers’ lower proficiency level and, therefore, they may exhibit avoidance strategies that prevent them from using deviant prosody. However, there are by-tune differences that apply in the case of the SpE learners, which mostly differ on a realizational level rather than a distributional one. Thus, a higher proficiency level in learner language may also lead to more potential prosodic deviances in natural spontaneous productions.

5.3 Discussion of intonational phrasing results

The interdependent prosodic patterns analyzed in the present study (tones/tunes, intonational phrasing, and f0 range) in non-native speech are further characterized, among other factors, by an interplay between fluency, proficiency, segmental features, syntax, and speaking style (see also Ford 1993; Levelt 1993: 36; Ford and Thompson 1996: 137; Gutiérrez Díez 2001, 2005, 2012; Gut 2009). The present study focused on the production of prosody by NNSs by adding an analysis of fluency features (filled pauses, silent pauses, hesitations), lexical complexity measures (TTR, Guiraud index), voice quality features (creaky voice, elongations of vowels and consonants), and the proficiency levels of the learners. The present section pulls together the results of intonational phrasing and explains the interplay between prosody and various other linguistic levels, i.e. fluency, segmentals, and syntax. Although not specifically focused on syntax (lexical complexity) and segments (besides the lengthening of segments), the present study makes assumptions about their interrelatedness with prosodic features. Overall, IU frequency, speech rate per IU (measured in WPS), and length (in seconds and words) in the present study has been found to be conditioned by the speaking style (monologues vs. dialogues), interview style (number of interruptions, interviewer style and speaker sex), language proficiency and fluency features (number of hesitations, filled and silent pauses, speech rate per IU, hesitations, and elongations), as well as regional variation (in the case of the SpE speakers).

5.3.1 Intonational phrasing in native speech

Before the results of the different interlanguages are discussed, it is important to discuss how NSs break up their utterances in the samples selected for the present study. While the NNSs differ considerably from the NSs in this study, there are also relatively high IQR values even among native speakers, indicating a high variability in the production of IUs, which is compatible with previous research (e.g. Gutiérrez Díez 2008: 341). Speakers are generally free in choosing when to break their utterances. However, intonational phrasing is not arbitrary and IU frequency, length, and speech rate is further restricted, among other factors, by the task/speaking style, possibly attitudes towards the interviewer's accent, and regional variation.

For instance, the speaking style determines how frequent prosodic breaks are and how long they usually will be. Dialogic speech produced by NSs has been shown to be characterized by fewer, shorter, and faster (speech rate) IUs in the present study. The NBrE speakers have been found to be interrupted frequently by the Belgian English female interviewer. In the present study, more interruptions have been shown to logically lead to more IP breaks. Despite the interruptions, there seems to be some evidence that IUs are planned holistically (Ladd 1986: 320; Levelt 1993: 400; Swerts and Geluykens 1994; Wichmann 2000; Clark and Fox Tree 2002: 94; Gilbert 2008: 10-11; Götz 2013). In narratives, IUs form the narrative actions (Chafe 1980: 69, 80), and they are turns in conversations (Clark 1994; Ford and Thompson 1996). According to Swerts and Geluykens (1994: 40) "speakers are also able to supply global prosodic structure in spontaneously produced speech", which "shows that, even when there is no predetermined structure, speakers are able to pre-plan their ongoing speech in such a way that its hierarchical global structure is reflected in its prosody". According to Clark and Fox Tree (2002), filled pauses (*um* and *uh*) and elongations of syllables have a similar function in speech, just like interjections such as *you know* and *I mean*. In addition, they are used for specific purposes in conversations, e.g. to signal that they are searching for a word, deciding the content of the next turn, or to hold the floor. According to them, while speakers plan their IUs, they are able to anticipate delays and choose one of the interjections to signal to the listener how long he/she has to wait to hear the next turn. They found that *uh* is used for shorter delays and *um* for longer ones, which seems to be consistent across varieties of English (AmE and BrE in this case). Interestingly, there is evidence that the duration of pauses in general correlates with how strong the prosodic boundaries are (Zellner 1994) and how long the phrases will be (e.g. Zvonik and Cummins 2003) (as summarized in Weingartová et al. 2014b). In the present study, the main difference between the SAmE speakers and NBrE speakers may lie in stylistic differences and the more frequent use of filled pauses and elongations by the SAmE speakers, because the SAmE samples in the present study mainly resemble narrative speaking styles that may need more filled pauses and elongations to signal delays to the listener and to add more effects to a specific part of their narrative. The NBrE speakers, on the other hand, may not need to plan their phrases too much ahead and may not need to rely on filled pauses and elongations as much, because they are involved in conversations with shorter question and answer sessions that may not require much planning of the IUs. Overall, SAmE females were found to produce significantly fewer IPs but more ips compared to the productions of the NBrE speakers. This may be explained by a more narrative speaking style of the SAmE female speakers and the need to break down their longer narrative into more manageable chunks (i.e. ips) for the listener and fewer long silent breaks that create longer and more coherent chunks (i.e. IPs). The reason why I think that this is a speaking style difference is because the AmEO group (which includes mostly dialogues) is not significantly different from the NBrE productions in any of the IU measures. Interestingly, however, the SAmE males do not differ from NBrE productions in IP frequency but they produce significantly fewer ips phw than the NBrE controls. This may point towards gender differences in narrative speaking styles, where male speakers produce longer IPs with fewer ip breaks.

The findings contrast with previous results reported in the literature on IU frequency and length of NNSs seems to be more characteristic of the NSs in my data set (e.g. Gutiérrez 2008: 341; Gut 2009: 228-229). Most NSs are below the IU frequencies and lengths reported in these two studies. This finding might be largely attributed to SAmE speakers' slower speech and NS-NS interactions (AmE corpora) and foreigner talk in NNS-NS interactions (LOCNEC). However, if IQR values are taken into consideration some speakers fall well within the results of previous research. Long (1983: 129) found that NSs produce shorter T-units (in words) when communicating with NNSs, in order to make the input more comprehensible, while Smith (2007) found that NSs of English do not always slow down their speech or produce shorter utterances when talking to non-natives. The NBrE samples were among the shortest and fastest IUs in terms of seconds and WPS. The IPs of AmE speakers exhibited the tendency to produce longer IPs. This result could possibly be explained by the fact that all AmE speakers were talking to other AmE NSs and did not need to slow down or break their utterances more often to ensure comprehension. The NBrE speakers were talking to a highly advanced Belgian English speaker with no audible accent, although she had a very emphatic high-pitched intonation. Thus, while the NBrE speakers do not slow down, they produce major IP breaks more often than the other NSs in the present data set. Therefore, one explanation for the tendencies in the present study may be that the native NBrE speakers engage in foreigner talk and produce shorter IPs. Maybe the NBrE speakers use the shorter IPs to compensate for their faster speech rate. The results may also point towards a cultural difference between AmE and BrE speakers. For instance, Tannen (1986) analyzed six American friends from different parts of the US. These sub-cultures made use of different conversational conventions, e.g. when to take over a speaker-turn: "no participant realized that the other's way of talking was a reaction to this or her own, but simply attributed it to the unaccountable and possibly negative personality or intentions of the other" (Tannen 1986: 209-210). This study shows that not only might cultural differences disturb L1-L2 conversations but also different sub-cultures in L1-L1 conversations. One of the impressions I have had while listening to and annotating the NBrE samples was that the Belgian English speaker interrupted the NBrE speakers at inappropriate times. This behavior of interrupting might have triggered more negative attitudes in the NBrE NSs, which, in turn, may have influenced their intonational phrasing behavior. The NBrE NSs might have considered the interrupting by the non-native interviewer to be ‘rude’ and, as a result, possibly answered in shorter phrases and less expressive intonation. For instance, speaker BE027 uses uptalk on the word *fridge* followed by a 0.74 seconds long pause in the following utterance: “you only have like one or two little freezer boxes in the fridge”. After the pause, the speaker continues with “we only got like …” and is then interrupted by the Belgian English female interviewer with “no, we have a big freezer here”. The rising intonation at the end of the utterance paired with the relatively longer silent pause may have been interpreted by the interviewer as a question or as an invitation to add to the conversation, even though the interviewee clearly intended to go on with his monologue. In sum, while the NBrE IUs may be characteristic for the speaking style (i.e. dialogues), there may also be features of foreigner talk present, which may partly be due to cultural differences in turn taking (Belgian speaker vs. native NBrE speakers).

According to previous research, AmE is slower than other varieties of English, and especially SAmE has been described to be slower, which is often attributed to differences in vowel quality and rhythm (Preston 1998: 145; Tottie 2002: 16; Robb et al. 2004; Jacewicz et al. 2009: 246; Baugh and Cable 2013: 365; Clopper and Smiljanic 2015). Some research, however, showed no difference in speech rate (e.g. Ray and Zahn 1990; Clopper and Smiljanic 2011: 241). The findings on intonational phrasing in native speech are consistent with existing research and confirms that on the IP level, all AmE varieties group more with learners who produce longer IPs in seconds and fewer words. NBrE clearly set themselves apart by producing the shortest duration of IPs (4.31 seconds (absolute number), 3.8 WPS). When the ip level is considered all native varieties are extremely similar. Generally, SAmE speakers in this study were found to produce more filled and silent pauses and both AmE groups used a lot more elongations than NBrE speakers. This might explain why especially the SAmE speakers are different from the NBrE speakers in IP length and IU speech rate, which is mainly due to pause and elongation frequency. The findings of the IU analysis show that compared to the NBrE data, the learners diverge immensely on all measures of intonational phrasing. However, when compared to other varieties of English (here different AmE dialects), different results are obtained, i.e. more similarities can be found. This suggests that NBrE might not always be the target model for all learners and the variety chosen might be one that is characterized by a particularly fast speech rate, making comparisons to learners difficult.

5.3.2 Intonational phrasing in non-native speech

The intonational phrasing of the three learner groups is generally characterized by longer IUs in seconds and fewer words (except CzE speakers), and a slower speech rate per IUs compared to native speech, however there are different groupings within the non-native group. The present section seeks to explain the different groupings by considering the interplay between fluency, segmental, lexico-grammatical features, and prosodic features of the different interlanguages. As Andreeva and Barry (2012: 262) nicely put it, “prosodic structure has been shown to operate within an intricate system of tonal and energy dynamics and of timing in relation to the segmental skeleton that carries it”. In other words, “[…] the entire F0 contour is specified at the phrasal level by means of a complex interplay between metrical structure, prosodic phrasing, syntax and pragmatics" (Arvaniti 2011: 1). How these results relate to previous research on the same learners will be discussed in the following.

One aspect that the present study focused on were selected fluency features, because previous research has shown that there is a correlation with fluency features (filled and silent pauses, speech rate) and prosody (Rasier and Hiligsmann 2007; Gut 2009, 2012a). While fluency studies on the same learners as in the present data set determined striking similarities between the speech rate of German and Czech learners (cf. Götz 2013; Gráf 2015), the present study found Czech learners to be among the ones with the longest IUs and fewest IU breaks. Generally, 1-2 word utterances were left out of the analysis because longer stretches of speech have shown to exhibit more deviances, that is why my study does not directly compare to previous studies who found shorter IPs to be the most frequent (and also because previous studies were based on the entire recordings and not 560 word samples). Taking the ip as the domain for comparison, the results are more comparable to previous studies (e.g. for NE speech Nevalainen 1992: 419; Cruttenden 1997: 72). Previous research has found that faster speakers produce fewer prosodic breaks and longer IUs (Trouvain and Grice 1999; Šimáčková et al. 2014). In the present study there was a tendency for faster speech in WPS to be produced the fewer ip breaks are made. However, the result was not statistically significant. Šimáčková et al. (2014) found that Czech-accented English, as well as AmE controls, in reading was characterized by fewer and longer IPs. They explained their results by the fluency and tempo of the more advanced learners, with accelerated tempo leading to a lower number of prosodic breaks and longer IPs. Perhaps the learners and maybe even some of the NSs were nervous about being recorded and they accelerated their speaking tempo and thus produced fewer breaks and longer IUs (in the case of the CzE speakers). Additionally, previous research has also found that fast speech might come with a side effect of an overall reduced f0 range and a simplification of tonal structure, e.g. monotonal tones instead of bitonal ones, but there is also considerable interspeaker variability (cf. Fougeron and Jun 1998; Trouvain and Grice 1999). Generally, bitonal tones were overall extremely rare in the present study, in native as well as non-native speech. Furthermore, previous research found that in non-native speech, faster speech might lead to other prosodic, pronunciation or other errors (e.g. Derwing and Munro 1997; Skehan and Foster 1997; Cauvin 2013). For this reason, an analysis of segmental elongation, creakiness, pausing behavior and hesitation phenomena was undertaken and combined with the analysis of intonational phrasing. Even though the CzE speakers stick out with the longest (in words and seconds) IUs and fewest IU breaks, their speech rate per IU is among the slowest. This is mostly the case because the CzE speakers produced the most filled pauses and elongations from all groups (but a native-like usage of silent pauses). Therefore, the use of filled pauses and elongations creates the impression of fluent and connected speech which glues the IUs together and therefore fewer ip breaks are produced in the case of the Czech learners. A mixed-effects model, however, determined that all learners produce a significantly slower speech rate in IUs compared to all NSs and that all fluency variables (filled and silent pauses, hesitations, and elongations) have a significant effect and lead to a slower speech rate in IUs.

At the outset of this study, it was hypothesized that differences in speaking style may also affect the intonational phrasing behavior in non-native speech. While the speaking style had an effect on IP frequency, length, and speech rate, with dialogic speech having more frequent IP breaks and faster speech rate in IPs in WPS, speaking style did not play a significant role for ip frequency and -speech rate. Thus, even though the samples are uneven as to their distribution of speaking styles by corpus (see Table 6 in Section 3.3), the learners all show the same trend: they all produce significantly slower speech rate in IUs. The number of interruptions also do not explain this result, because the NBrE speakers are interrupted by the interviewer as frequently as the CzE and GerE speakers in this sample, while both AmE groups and both Spanish groups are interrupted only half as much. Therefore, the number of interruptions cannot explain the similar trend in IU speech rate. However, when it comes to IU frequency the GerE females clearly stick out with more IPs phw. Interestingly, if the findings of the present study are compared to findings from previous research on read speech (Mennen et al. 2014), the GerE speakers produced a median of 6.2 IPs phw (compared to 6.9 IPs in read speech) and the NBrE speakers produced a median of 5.4 IPs phw (compared to 5.7 IPs in read speech), while their IP length in seconds is quite similar (5.02 seconds (median) in GerE and 4.31 seconds in NBrE) (see Mennen et al. 2014 for similar findings in read speech). Therefore IP frequency may point towards an L1 influence in the case of the GerE speakers, which also seem to yield similar results to read speech. The point this paragraph was trying to make, is that NNSs only show a similar behavior in speech rate per IU (with significantly slower speech rates per IU) but show L1-specific differences when it comes to IU frequency, which may stem from speaking style differences as well as their proficiency level (and fluency level, more specifically). L1 influence and proficiency will be discussed in more detail in the next section. The speakers from Murcia produce the lowest number of words per ip, a slow speech rate per IU, as well as the fewest IP breaks (male speakers). One reason for that may be that Murcia speakers have a very monotonous speaking style and thus no prosodic boundaries could be annotated/were not perceived by annotators (see also Swerts and Geluykens 1994; Pickering 2004: 31-32). Since the present study focused on prosodic boundaries, this may have been obscured in some of the learners with lower proficiency levels, who rely more on syntactic marking of boundaries instead of prosody. However, considering pausing behavior has added a level of comparison between these intonational phrasing, which will be explained in the next section.

5.3.3 L1 influence or developmental factors in intonational phrasing

Previous research on intonational phrasing, although scarce, found Czech, German, Spanish, and English in direct comparison to produce similar number and length of IU in comparable contexts (Gutiérrez Díez (2012: 220) Mennen et al. (2014: 316) Chamonikolasová 2017: 46). However, Mennen et al. (2014: 316) found German speakers to produce more IPs than English NSs in read speech (ca. one IP more on average). Therefore, I hypothesized that learners will not exhibit any L1 interference problems but rather developmental issues should be at play. IUs were selected as the basic units of analysis in the present study, because they are the bearers of all other prosodic features. Moreover, previous research has shown that IUs are well suited for the analysis of the level of automatization and proficiency level of learners (Lintunen et al. 2015). An intermediate learner with a lower proficiency level, especially in spontaneous speech, might be more occupied by lexical retrieval (Gut 2017: 218), producing segments correctly, and with word order, rather than paying attention to the pragmatic effect his/her prosody might have, which also has an effect on their L2 fluency production and ultimately influences intonational phrasing. This tendency of learners to focus on lexis might stem from L2 instruction in general focusing mostly on lexico-grammar instead of including prosody as well. In this regard, Tahta et al. (1981b: 371) speak of a trade-off effect, where one linguistic level suffers at the expense of another: “We know that younger learners have more trouble with syntax than other learners, so that in the first year or so of acquisition proper pronunciation is more likely to be sacrificed until basic mastery of the L2 has been achieved […]” (Tahta et al. 1981b: 371). Similarly, Nava (2008: 163) also suggests that learners rely more on syntax than on stress until they have reached an appropriate level of syntax. Generally, one can assume that the lower the proficiency, the lower the fluency and accuracy, and consequently the shorter IUs will be in words and longer in seconds, the more accents are produced, and the more f0 resetting occurs. With increasing proficiency, the learners should show gains in their lexical repertoire and require less words to circumscribe certain words. This was investigated more closely in the present study with the CEFR levels, fluency measures, and information on stays abroad. Higher proficiency is posited to contribute to the superior performance of advanced learners over lower level learners when it comes to the segmentation of intonational phrasing. Jiránkova et al. (2019: 33) found speech rate to increase in L2 English from C1 - C2 levels, which also confirms the findings in the present study. The present study investigated learners of English at different levels of proficiency (45 CzE speakers and 19 GerE speakers) and found there to be a negative and weak correlation of IU frequency with increasing proficiency, i.e. the number of IUs decreases with increasing proficiency. In the case of the CzE learners, this leads to more non-native-like intonational phrasing patterns and more native-like patterns for the GerE speakers. However, the more years of English the CzE learners have had at university, the higher and thus more native-like their ip frequency becomes. The fact that CzE speakers improve their ip frequency with increasing years of English at university may support Gut’s (2009) findings that English language instruction (although it may not be focused on prosodic instruction *per se*) may have a bigger influence than a stay abroad experience in an English-speaking country. Additionally, IU speech rate was positively correlated with increasing proficiency, with higher CEFR levels leading to a faster speech rate within IUs. However, the correlations found between intonational phrasing patterns (frequency (weak correlation) and speech rate (moderate correlation)) always had a stronger effect for the CzE speakers, probably due to the larger amount of CEFR information available for these learners. IU frequency and speech rate for all NNSs shows no improvement after a stay abroad. In sum, while the learners’ stay abroad experience does not seem to dictate their IU frequency and speech rate, the current developmental stage in second language acquisition (CEFR levels (GerE for both IU frequency and speech rate and CzE for IU speech rate only) and YOEAU (CzE only for IU frequency)) seems to point towards improvement of intonational phrasing that both CzE and GerE speakers share. However, it has to be kept in mind that all correlations mentioned here were weak to moderate, based on imbalanced data available on the learners, as well as a pseudo-longitudinal analysis and, therefore, have to be interpreted with caution.

Besides providing insights into a learner’s current developmental stage, IU speech rate may also be evidence for L1 influence. Gráf (2019) found that NSs (LOCNEC) produce the fastest speech rates measured in wpm. Overall, the fastest speech rates were reached in the dialogic speech form, although he found the difference between dialogues and monologues to be small. However, the picture description sticks out with considerably slower speaking rates in native and non-native speech. This confirms the findings of the present study where NBrE speakers were found to be among the fastest speakers and generally there were no larger differences in speech rate in dialogues and monologues (just IPs were generally produced faster in dialogues but not ips). This is not surprising considering Gráf (2019: 184) also confirms my finding that many monologues are, in fact, more like conversations because of a lower proficiency or sustained thinking breaks, which lead to several interventions by the interviewer. In the present study, I have deviated from the LINDSEI conventions of labeling speech from Task 1 as monologic and Task 2 as dialogic, but I created my own labels by inspecting each sample closely. Jiránkova et al. (2019), however, find contradicting results by using syllables per minute (spm) instead of wpm: "If speech rate were measured in syllables, Czech would be faster than English but identical or slower if it were measured in words". If one inspects the results presented by Gráf (2019: 180), it becomes clear that the same L1 Czech speakers, who speak in their L1 in tasks 1 and 2 in the LINDSEI-CZ supplement, produce a slower speech rate in their L1 than the NE speakers. Even if standard deviations are taken into account, they do not reach the 202 wpm mean in Task 1 (151 wpm in NNE and 157 wpm in their L1) and 208 wpm in Task 2 (156 wpm in NNE and 158 wpm in their L1). Therefore, in both tasks, the Czech speakers are slower than English speakers by over 40 wpm, no matter if they speak in their L1 or their L2. This may be interpreted as L1 influence that Czech(-accented) speech in general is slower than NBrE speech. However, one has to consider that the Czech speakers are highly proficient bilinguals, which is a fact that can alter their L1 speech rate, since previous intonation research has shown that L1 attrition can take place (De Leeuw et al. 2012; De Leeuw 2019). Therefore, a reduced L1 use may possible lead to slower L1 productions (Jiránkova et al. 2019: 29). In their study, the Czech speakers produced 276.26 spm in Task 1 (compared to 202.61 spm in their L2), and 277.13 spm in Task 2 (compared to 208.39 spm in their L2). In sum, Jiránkova et al. (2019) found that CzE speakers tend to transfer their speech rate from their L1 into L2 and speakers with a slower speech rate in their L1 are also very likely to be slower speakers in their L2. The present study, however, also found that IU speech rate is conditioned by several fluency phenomena. Despite Jiřelová's (2018) study being based on only eight speakers who were recorded in their L2 (LINDSEI-CZ) and L1, a tendency for speakers retaining their speech rate, filled pauses, and repeats was observed, i.e. the slower their speech rate, the more filled pauses and repeats they use in the L1, the more these features proportionally increase in their L2. Therefore, suggesting evidence for L1 transfer for all three fluency phenomena. Jiřelová (2018) also proposes that a less varied use of discourse markers in the L2 of the Czech speakers could explain the higher use of filled pauses. Additionally, Jiřelová (2018) found in a pilot study a very frequent vowel prolongation when Czech L1 speakers pronounce the conjunction, *a* (and) in Czech (32). In the present study, the CzE speakers have been found to produce the most phrase-final elongations and filled pauses, as well as the slowest speech rate within ips, which could be taken as instances of L1 transfer (cf. Dankovičová 1998; Volín and Skarnitzl 2007; Jiřelová's 2018). In sum, while in my reduced sample of the selected corpora there seem to be minor speaking style differences at play, L1 influence of segmental and fluency phenomena may also account for the variation found in speech rate per IU (measured in WPS) in the CzE samples.

As described earlier, non-native speech in this study was characterized by longer (in seconds but not words) and slower speech in WPS in IUs. The fact that all learner groups have most of these characteristics in common, regardless of their L1 background, seems to suggest that developmental issues are at play. However, GerE females produced more IP breaks and CzE female speakers produced fewer IU breaks than NSs. Learners may be compensating for their lower lexico-grammatical knowledge with different strategies, i.e. lengthier explanations, more frequent self-repairs, and different pausing behavior, which, in turn, leads to more frequent and longer intonational phrases and possible other prosodic phenomena, i.e. f0 resetting (change in pitch level) and tonal choices (for instance, more high tones to connect ips). As discussed in the paragraph above, "often language learners take more words and more attempts to convey what they want to say" (Ward and Gallardo 2017: 18). Similarly, speakers in Mauranen's (2006: 138) study seem to use lengthy self-repairs, in which the ELF speakers rephrase their own speech, while the NSs do not use self-repairs often. However, this does not seem to be the case for most learners in the present study, while they produce longer IUs in seconds, they are not longer in terms of words. However, if measured in WPS, learner language is slower. The issue, therefore, might be rather temporally-based, fluency-related, i.e. learners produce more pauses and may produce longer syllables and/or more silent pauses. While NSs have been found to hesitate without producing extra accents, NNSs have been found to produce contextually inadequate accent patterns before and after speech pauses (cf. Rasier and Hiligsmann 2007: 58, with L1/L2 Dutch and French speakers). What may explain the more frequent ip breaks is the proficiency and accuracy of the speakers, which influences their fluent production of speech. If the speech flow is disrupted by dysfluencies and hesitations/repetitions, then this leads to frequent restarts of an utterance as well as the resetting of f0. Thus, the lower the accuracy, the lower the fluency and the more IUs are produced. In addition to lower proficiency in an L1, which may cause these dysfluencies and more frequent IU breaks, L1 influence may be at play. Romero-Trillo and Llinares-García (2004: 75) found SpE speakers to make use of the opposite mechanisms to mark prominence compared to NE speech, i.e. they use duration to signal prominence in unmarked utterances and produce marked elements with amplitude. The NSs did exactly the opposite. They explained their results by a possible L1 influence because Spanish is a syllable-timed language. Spanish learners often transfer fillers from their L1, where a pause is inserted at the end of utterances to focalize words, which contributes to the perception of non-nativeness in speech (Clark and Fox Tree 2002: 93; Ortega-Llebaria and Colantoni 2014). The SpE speakers, which have been found to use elongations frequently in their L1 and L2 (García Lecumberri 2017: 184), were not found to use elongations more frequently than the other learner groups. Inspecting the differences in the investigated fluency features more closely for the MurE and MadE groups, shows that the speakers' proficiency levels may be more revealing than L1 influence, since different trends can be observed in the data for these two groups. MadE speakers produce native-like speech in the use of filled and silent pauses, as well as elongations, while MurE produce more filled and silent pauses but not elongations. Both produce more hesitations the most of all learner groups but the deviance is more extreme in MurE productions. For speech rate for both IU levels, MadE and MurE can be found at opposing poles, i.e. MadE producing the slowest IU speech rates and MurE the fastest of all learner groups (fastest for ip level but the same as GerE speakers on IP level), all other factors being accounted for. This, is a surprising finding, given that the fluency of MurE speakers has been determined to be less native-like but still their speech rate is the fastest among all learners. Further studies would be needed to analyze the syllable length in the MadE and MurE samples to see whether an L1 influence can be detected in the production of IUs. In sum, all learners create extra pauses and other dysfluencies, which leads to additional IU breaks and longer IUs and this may be conditioned by the current proficiency level and/or L1 transfer. These examples demonstrate that the production of IUs is conditioned (among other factors) by the level of mastery of lexico-grammatical units, markedness, and possibly L1 interference. The L1 transfer aspect will be discussed in more detail in the next paragraph.

Research on segmental features found that learners with syllable-timed L1s have difficulties acquiring grammatical words in their weak forms (i.e. temporal reduction) and thus pronounce these words as significantly longer (Volín and Johaníková 2018: 182), CzE stressed syllables tend to be shorter and unstressed syllables longer than in NE speech (cf. Weingartová et al. 2014a). Thus, if learners have not yet acquired the L2 syntax and segments properly, deviations in intonational phrasing may be expected. In addition to pausing and hesitation phenomena frequently produced by NNSs, the developmental process and lower proficiency may account for the significantly longer and slower IUs compared to native speech. The results on the lack of temporal reduction of grammatical forms might explain why the Czech learners produce by far the longest ips (median length=2.43 seconds). The CzE speakers are very advanced speakers of English but might not have acquired the weakening of grammatical words yet. This would have to be investigated in future research. Future studies will also have to determine why the Czech learners behave differently from Spanish and German learners in their prosodic phrasing, i.e. IU frequency (not speech rate per IU). One explanation may be L1 influence and the delimitation of APs instead of IUs. Indeed, in the present study, it was found that the CzE speakers use many strategies of connecting and separating IUs from each other and, therefore, create an impression of highly fluent speech. The CzE speakers are among the speakers with the most elongations and the fewest IU breaks. This may suggest an L1 influence, since Czech intonation has been shown to typically use phrase-final lengthening to mark prosodic boundaries (cf. (Dankovičová 1998; Volín and Skarnitzl 2007). In comparison, Chamonikolasová (2017: 46) found the tone unit length in Czech to be 1-14 words long and 1-19 words long in English. The present study found the following normalized word ranges (phw) for each group: NBrE 1-17, SAmE 1-34, AmEO 1-13, CzE 1-8, GerE 1-25, MadE 1-10, MurE 1-8 words. These results show that native speech includes longer ips and non-native speech shorter ips in words (except the GerE speakers). However, the SAmE and GerE speakers produce extremely long ips. Furthermore, the average (median) word count by ip is very similar to Chamonikolasová (2017: 49) study, who found 4.25 words per tone unit in Czech non-scripted dialogues and 4.20 in English ones. The speech rates (WPS) by group per ip are the following: NBrE 4.17 (1.8 IQR), SAmE 3.83 (1.8 IQR), AmEO 4.07 (1.7), CzE 2.89 (1.2 IQR), GerE 3.12 (1.3 IQR), MadE 3.14 (1.4 IQR), and MurE 2.78 (1.4 IQR). This result shows that generally the NSs have a higher word count by ip, compared to all NNSs, with the NBrE speakers producing the most words per ip. The fact that the two SpE groups have extremely similar results when it comes to ip length in words and seconds shows that this is a feature that may point towards L1 influence. In comparison, Gutiérrez (2001) found the English learners from Murcia in a reading task to produce considerably longer stressed and unstressed syllables compared to NSs of English, which contributes to a slower speech tempo. Gutiérrez sees this as a developmental error, likely to be caused by lower proficiency level and “in the application of canonical articulatory speed” (2001: 108). Maybe this can be seen as support for a more syllable-timed rhythm in SpE and, thus, evidence for L1 influence, which was also found in previous research (e.g. Gut 2009; García Lecumberri et al. 2017). In the theory, German and English have been described as "demarcating languages", which tend to produce more IPs than "grouping languages" such as most Romance languages, which tend to group long strings into fewer larger constituents (Lleó and Vogel 2004: 80). While the female GerE speakers exhibited the trend to produce more IPs (demarcating function), CzE females and MurE males tended to produce fewer IU breaks (grouping function). These trends, however, are rather unlikely to be due to an L1 influence, since gender-based differences can be found (only CzE and GerE females and MurE males exhibit this trend) as well as regional differences (while MurE males exhibit this trend, MadE speakers do not). In sum, segmental and fluency features seem to be responsible for some of the variation in intonational phrasing in non-native speech, which can be explained by a possible interplay of proficiency level, L1 influence, and gender-based differences.

To address RQ1, on the structural level, the learners' speech is characterized by longer IUs (in seconds but not words) and slower speech in WPS, while IU frequency depends on the L1 variety spoken and the learners' gender (i.e. females being more deviant). To address RQ2, it can now be said that while all learners produce slower speech within IUs, only GerE female speakers produce significantly more IPs, CzE female speakers produced significantly fewer IU breaks, while the SpE speakers adopt language-appropriate values with regard to IU frequency (except MurE males for the IP level) (RQ3a). The GerE females are also the ones with the most high edge tones, and this might suggest a need to split up their utterances into smaller and more manageable chunks for the listener and to "glue" these together with the use of high edge tones (RQ3b). These results suggest that there are universal (on the realizational level of IU speech rate), L1 variety-specific, and gender-specific differences (on the distributional level of IU frequency) in the intonational phrasing of advanced spoken interlanguage that deviate from NE speech. Generally, female NNSs may feel more of the need to make themselves understood and may employ non-native-like strategies to break up their IUs into smaller units, accompanied by the use of high edge tones. The CzE female speakers produce highly fluent speech, but at the same time make use of the most filled pauses and elongations and rank second in the production of high edge tones, which glue the utterances together and create almost endless IUs. The GerE females employ other strategies. The results, even though not all of them proved to be statistically significant, point to the fact that all learners from the three different L1 backgrounds show similar developmental paths, which are reflected in their IU speech rate and length but not in their IU frequency. For IU frequency different L1-specific and gendered strategies of chunking are used, which are also conditioned by a speaker's proficiency/fluency level.

5.3.4 Summary

Section 5.3 has explained the interplay between prosody and various other linguistic levels such as fluency, accuracy, syntax (lexico-grammar), and segments. Variation (MadE vs. MurE; NBrE vs. S/AmE/O) has to be accounted for because intonational phrasing (IU frequency) has been shown to differ by region. The difference in IP length and speech rate between SAmE and NBrE speakers was determined to be mainly caused by phonation and pausation time, i.e. tempo. On the IP level, fluency (filled and silent pauses) and segmental features (vowel quality and lengthening) contribute to differences in length (in words and seconds) and speech rate (in WPS). The ip level, where fewer pauses are used within these IUs and which correspond to most IUs used in previous research, show the differences between native and non-native speech more clearly. Thus, in sum, regional variation (native and non-native), gender-based differences, proficiency levels, L1 influence, and different intonational phrase levels have to be accounted for to describe the prosodic behavior of non-native speech. In my opinion, the major factor that determines the prosodic structure selected by a learner is a combination of L2 proficiency and L1 influence. While speaking style played a minor role in IU production in native speech, it played an even marginally role in non-native speech. Pseudo-longitudinal intra- and inter-group analyses uncovered different effects of the proficiency-related variables. The stay abroad variable did not show any improvement towards more native-like intonational phrasing in CzE, GerE, and MadE speakers. However, a common developmental path for improvement with increasing proficiency was observed for CzE and GerE speakers. Thus, highlighting that different proficiency parameters need to be analyzed together. The current proficiency level was determined to be the more likely explanation for the IU patterns, because all learners exhibited similar deviations from the NS norm and not all L1 influence were observed. There seemed to be stronger L1 effects for CzE speakers than SpE speakers, and GerE speakers exhibited the least L1 influence. Additionally, learner syntax is another feature that needs to be addressed in future research when analyzing L2 phrasing proficiency. What this means is that linguistic levels cannot be viewed and analyzed independently of each other if the goal is to get a complete picture of L2 speech. Pragmatics is another area that is rarely considered together with L2 prosody. While the pragmatic functions had only a minor effect on intonational phrasing, i.e. finality and handing over of turn producing faster speech in WPS, while listing and insecurity elicit slower speech in WPS, the effect of pragmatic functions on the tonal choices made was expected to be more significant. Therefore, the pragmatics of prosody will be discussed in more detail with regards to the findings of tonal choices in Section 5.5 (see also Section 5.2 for such a discussion).

5.4 Discussion of f0 range results

As was discussed in the theory and results section, F0 range is a gender-, culture-, and discourse-specific phenomenon. There is a wide range of factors that influence the shape of f0 range in native and non-native speech, e.g. vocal fatigue, time of the day, ambient noise, the speaking style, the emotional content of the dialogue, the degree of familiarity between the interlocutors, etc. (cf. Braun 1995; Levitan 2014). While gender has been found to consistently influence f0 range values in the present data, i.e. women have higher levels and wider spans than male speakers, discourse-specific phenomena were also observed. One discourse-specific phenomenon analyzed in the present study was prosodic entrainment. In sum, it was determined that f0 range is a complex interaction of gender, speaking style, and social distance, among other factors, and the findings will be discussed in light of the research questions in the following sections.

5.4.1 F0 range in native speech

The native corpora in the present study included different dialects of English, Northern British English and four different American English dialects: Northern, Western, Midland, and Southern. Each of these dialects can exhibit typical f0 range patterns for their region. As in previous research, the present study found AmE speakers to be among the speakers with the most extreme highs and lows as compared to NBrE speakers (e.g. Feagin 1997: 134; Shevchenko and Skopintseva 2004; Jilka 2007: 85; Barry 2008: 115-116; Fox et al. 2013: 123). The NBrE prosody in the present study can be described as rather flat but with a typical declination trend on declaratives (HL-tune). As far as f0 level is concerned, SAmE speakers produced a significantly higher f0 level for HL-tunes and significantly lower LL-tunes when compared to NBrE speakers. Also for F0 span, SAmE and AmEO speakers produced a significantly narrower f0 span in LL-tunes. Besides these difference, the fact that two AmE groups display larger similarities to each other in their f0 range than to the NBrE group provides evidence for viewing them as part of one dialectal group, i.e. GA English. In addition to that, just as in non-native speech, NSs vary significantly in their f0 range by sex, and there is also considerable individual difference. Additionally, cultural differences could be at play and wider f0 range could be attributed to speakers signaling their gender identity. Cultures of languages have their particular 'vocal image', reflecting socio-culturally desired personal attributes and social roles, and speakers choose a pitch that approximates the vocal image they want to project. Thus, f0 contributes to the perception of character types, such as, the wider their f0 range, the more positively speakers are characterized. How the speakers intended to portray themselves is impossible to find out in the present study, but clear dialectal influences seem to be present in the f0 range results.

One big factor besides dialectal, individual, and gender differences in f0 range is entrainment. According to the CAT (Giles et al. 1991), women entrain more than men because of a power imbalance that leads the less dominant speaker to entrain more. Romero-Trillo (2019) found that NSs (the interviewers) in LINDSEI-SP used more emphatic-rises and, thus, engage in foreigner-talk. However, this type of rise on backchanneling was not found as frequently in L1-L1 interactions in the LLC corpus. Nevertheless, the LINDSEI-SP corpus also includes an AmE female interviewer. Within interactions of AmE speakers, females seem to make greater efforts to keep the conversation going, they provide more feedback and support, they converge more to their interlocutors, and they ask more questions and confirmation requests than male AmE speakers (reviewed in Wolfson 1989: 176-182; Levitan 2014). In sum, previous research has found that there is a gendered and cultural difference which can influence the use of intonation in interactions. Also, in the present study, it was found that AmE native interviewers used a more emphatic intonation, which might have been adopted by them by engaging with NNSs. The AmE speakers from the NWSP corpus are interviewed by a lecturer, which results in the recordings sounding more like formal interviews than more natural dialogues, while some speakers from the NSV are interviewed by peers and, thus, exhibit a more relaxed and informal speaking style. The SAmE speakers tended to entrain their f0 less to that of their interviewers (result n.s.), which is mainly due to the low amount of talk by the interviewers because of a general monologic and narrative speaking style in the SAmE recordings. There seem to be quite considerable differences in style between the native samples. In the NBrE recordings, there is only one interviewer, a non-native female with an extremely emphatic intonation. Generally, the picture emerges of the male interviewees to be less interested in the interview, which is reflected in their short utterances and relatively flat intonation and lower entrainment values (DID-scores). The lower entrainment values may also be due to larger differences between the f0 range of the male interviewees and female and emphatic interviewer. In sum, speaker attitudes and foreigner talk, but mainly regional differences can explain the f0 range results for the NSs.

While this seemingly flat and monotonous intonation produced by the NBrE speakers may be attributed to disinterest in the interview/er, it is more likely that this a normal f0 range for the original accent spoken in Lancaster. While no prosodic studies on Lancaster English have been carried out before, Liverpool English may be a variety that might come geographically seen closest to the Lancaster speakers (LOCNEC). However, Liverpool English has been described to use rising tones at the end of utterances with statements, i.e. using uptalk and having a small overall f0 range (Knowles 1973: 180). While Nance et al. (2018: 286) found Liverpool English speakers to have a narrow f0 range (1-2 STs), they uncovered no evidence to suggest that HRTs are used in declaratives in Liverpool English in read speech. In Nance et al.'s (2018) study, four out of nine speakers spent time at the university of Lancaster. While L* L-H% non-uptalking tunes were the most frequent for the Liverpool speakers, in my data set H* L-L% tunes were most commonly used, suggesting that Lancaster English may be more similar to SBrE instead. Nance et al. (2018: 287) also found individual variation within their speakers and some Liverpool speakers resembling SBrE by using more H* L-L% tunes and wider f0 ranges. In my data set, no evidence that uptalk is used was found either in Lancaster English. While one can clearly hear in the recordings that high tones at the end of declaratives is not part of the linguistic repertoire of the Lancaster speakers, they also have a quite monotonous f0 range compared to the two AmE groups. The NSs (all male speakers and SAmE females) also use low-flat tunes often, almost as much as the NNSs. Generally, the more flat tunes in both low and high-registers seem to be more common in male speech than in female speech. The impression emerged that some of the native NBrE speakers had a negative attitude towards the interviewer or disinterest in the task, while the AmE speakers (especially NSV corpus) were engaged more in a narrative mode. The f0 range characteristics of narrations have already been discussed in the theory (Section 2.7.1) and may explain some of the variation in f0 range between AmE and NBrE speakers. Thus, the wider f0 range by AmE speakers may be partly explained by the story telling style of, for instance, childhood experiences that are often dramatic with twists and turns that require livelier f0 ranges than talking about university small talk (e.g. books, movies, traveling, etc.), as is the case for the NBrE speakers.

In sum, the NS groups exhibit similar f0 spans but both AmE groups produce narrower f0 spans for LL-tunes compared to NBrE speakers, and SAmE speakers produce higher f0 levels. All of this can be explained mainly by regional differences, sex differences, and differences in speaking style. However, the differences between the native groups in f0 range are rather small and the NNSs deviate considerably from all three NS groups, as is explained in the next section.

5.4.2 F0 range in non-native speech

My prediction was that learners would produce tunes closer to their L1 productions and that the production of native-like f0 range would be equally challenging for all learners, because all L1s (Czech, German, and Spanish) were found to exhibit a narrower f0 span than English. The results of the present study are consistent with previous results that, irrespective of the L1, the f0 span is always narrower in L2 productions and it is task-dependent (cf. Abu-Al-Makarem and Petrosino 2007; Mennen et al. 2007; Zimmerer et al. 2014). Zimmerer et al. (2014) found that German and French L2 learners of French/German respectively, produce a smaller f0 range in their L2s, suggesting that the main explanation for this variation is that the learners are less confident in their L2 productions and focus more on the production of segments and words (BV; Klein and Perdue 1997). The use of rising intonation and compressed f0 range is often explained by the learners supposed uncertainty, lack of confidence, or moderate anxiety (e.g. Mennen 1998: 18; Zimmerer et al. 2014: 1037; Volín et al. 2015: 121, Volín et al. 2017: 62). An alternative explanation may be that f0 can rise in stressful situations for some speakers (Streeter et al. 1983: 1359; Hollien 1990: 259). These studies, however, showed that acoustic and prosodic correlates of stress are not always good indicators of the actual stress level a person experiences, i.e. while some speakers may succeed at 'hiding' their stress level, for others this might be more visible. In many cases, the learners reported a high level of nervousness that may have affected their performance (there was a tendency for higher f0 levels for all learners for the most common HL-tune, though not significant). While stress is certainly a factor that could explain some of the results on f0 range in the present study, it is impossible to measure the level of actual stress the speakers experienced during the recording. Therefore, other factors than stress and uncertainty that explain the f0 range results will be discussed in the following.

The present study found in general that SpE speech is one of the narrowest in terms of f0 span among the learners for HL-tunes. The LF-tunes with an extremely narrow f0 range were frequently used in the Madrid corpus, and by German males. Interestingly, however, the MurE speakers produced significantly higher f0 levels and wider f0 spans for all other tunes (HH-, LH-, and LL-tunes). While I suggested in Section 5.2.3 that the GerE and CzE speakers may find themselves in a prosodic exploration phase, the MurE speakers may actually also try to explore their prosody by overhitting all other tunes except the HL-tune. Generally, however, in the most common tune (HL-tune), SpE speakers are the groups with the narrowest f0 span. Romero-Trillo and Newell (2012) analyzed the length and pitch of pragmatic markers as feedback (*mhm* and *yeah*) in the LINDSEI-SP and found that the initial and final pitch level of NNSs is lower and the pragmatic markers are shorter in their duration than the native pitch level, which makes them sound more assertive in their feedback than NSs.

> The result of the analysis shows that the combination of longer markers with the higher final pitch frequency constitutes the reason why these elements behave more interactionally in the native speakers, compared to the somehow more transactional flavor of the markers in the non-natives. (Romero-Trillo and Newell 2012: 129)

This extremely narrow f0 span could be interpreted to be typical of non-native speech, since they have a higher planning pressure and might be focusing too much on what to say and not on their intonation. According to Costa et al. (2008: 540), learners (especially lower-proficiency learners) have a an extra attentional demand and processing load which reduces the overall automaticity of the conversation. In sum, while all learners deviate in the most common tune by producing a narrow f0 span, they produce significantly wider f0 spans in HH- and LH-tunes and MurE speakers additionally produce a higher and wider f0 in LL-tunes as well. These differences were already hypothesized to be due to the learners' need to make their pragmatic intent overly explicit with high-register tunes.

A combination of speaking style and social variables (gender, familiarity, etc.) seem to offer another explanation of the f0 range results, mostly concerning those learners who do not deviate much from NSs. Some of the interview-pairings required a more active role for the interviewer in conducting the interview. In LOCNEC, the interviewer was a lot more active than the interviewers from the LINDSEI sub-corpora and, thus, many of the interviews in the monologues rather resemble lively dialogues. Therefore, the more dialogic the speech, the more entrainment was expected to take place. Entrainment was expected to be more robust in the L2 interactions with either native or non-native interviewer, because of a stronger need to enhance communicative success. The learners are at the university and may see this as a language test situation, that is, they may want to show off their English language skills to their L1 or L2 interviewers. Therefore, entrainment was expected to be stronger than in L1-L1 interactions. Generally, L1-L2 speaker interactions have been shown to often have the same impetus as the L1 speakers, who mostly control interactions with NNSs, which may or may not be maximally beneficial to the non-native listener/interlocutor (Zuengler 1991; Smith 2007; Lewandowski 2012). However, convergence can go both ways. Smith (2007) showed that NSs can choose which phonetic features of their speech are used to accommodate to non-native speech. In Lewandowski's (2012) study, the AmE and BrE NSs also showed a convergence in their speech to the German non-native subjects. L2-L2 interactions, on the other hand, have been observed to have an increased effort towards communicative efficacy and NNSs select features of their pronunciation which might potentially be unintelligible for the non-native interlocutors (Jenkins 2002: 91). However, this depends on whether the non-native interlocutor shares the same L1 or not (cf. Jenkins 2002: 93). Just as in native speech, learners will only be able to adjust their pronunciation if the features are within their phonetic and phonological repertoires (cf. Jenkins 2002: 95). That is a reason why it is important to know exactly what the relationship status and level of familiarity of speakers is. Because if the interviewees in my data knew that the interviewer speaks his/her native language proficiently, this may have an effect on the direction of entrainment. All interviewers in my data were familiar with the non-native accents of the students they interviewed. Living and working in the country makes the interviewers more likely to be familiar with the learners L1s. Thus, the native interlocutors are likely to have a very positive attitude towards the accents of the learners. In sum, L2-L2 interactions may be the ones where most entrainment takes place. This was, however, not testable in the present data, because the only L2 interviewer in the learner data was the male Czech interviewer in LINDSEI-CZ and, all other interviewers were NSs of the respective varieties of English. The present study investigated how tone position and speaker status (native vs. non-native) factor into accommodation and found that entrainment is higher in earlier parts of an ip and is reduced in later parts of an ip, in both native and non-native speech with NBrE speakers entraining the most and SpE speakers entraining the least. However, previous research has found that the entrainment of f0 range is an optional feature and even misalignment does not lead to failure of comprehension (Kim et al. 2011: 141). Overall, however, all groups (except the SpE speakers) entrained equally to their respective interviewers (male/female and native/non-native). Therefore, f0 entrainment does not seem to pose a problem to the CzE and GerE speakers. Why the SpE speakers tend to behave differently with regard to entrainment is explained in the next paragraph.

Previous research may also suggest that modality-specific effects may be at play in how f0 range is used to entrain to interlocutors. In Zellers and Schweitzer's (2017: 2338) study, speakers employed modality-specific prosodic strategies, i.e. they found that convergence was stronger when participants did not see their interlocutor than in face-to-face conversations. Thus, when access to facial expression, eye and body movement, gestures, etc. is given, participants seem to have less of a need to be "prosodically expressive" (Zellers and Schweitzer 2017: 2339). It is possible that the learners also used gestures to signal uncertainty in the negotiation of the meaning of their utterances or other pragmatic meanings (e.g. handing over turn, holding the floor, etc.) and were therefore (possibly) less prosodically expressive (in the case of HL-tunes). Perhaps there is also a cultural difference by the Spanish learners, who may rely more on body language and facial expressions than learners with a Germanic or Slavic language background. One reason for such a hypothesis is that the learners from the Spanish corpus were often found to use a clicking of the tongue, in order to signal the need for more time to think how to formulate the next utterance or that they cannot think of a word at the moment. It is very likely that the clicking of the tongue was accompanied by facial expressions and body movement to signal the need for more planning time. Therefore, future studies may benefit from "[e]xamining prosody and non-verbal visuals in conversation [which] provides a more holistic picture in understanding implied meaning" (Culpeper 2018: 161). In sum, the narrower f0 span used in HL-tunes may be partially explained by the use of gestures, hand movements, and or facial expressions, which may be used differently by the learner groups, possibly due to their L1 background.

The present study also confirmed previous results that rise spans are generally wider in final position (Barry 2008: 153, with SoCal and London female speakers). Why speakers produce larger f0 ranges in turn-final than in turn-medial phrases could possibly be to signal that they intend to continue speaking and do not wish to be disturbed in the middle of their narrative account. Just as in the use of filled pauses that signal the anticipation of a longer pause (cf. Clark and Fox Tree 2002), the larger the rise at the end, the longer the interlocutor may anticipate to wait and the more tolerant the listener will be in allowing a longer silent break before intervening. Thus, the learners in the present study have also demonstrated that they make use of f0 range to signal certain pragmatic functions, i.e. whether they intend to hold the floor or to cede the turn, to begin a new unit, etc. The learners in the present data also showed that they are able to set quoted speech apart by using more 'exaggerated' prosody. Tyler (2014) argues that if a marked and exaggerated prosody, which is unusual for the speaker's default f0 range, even if they quote themselves, may be a case of speech stylization (see also Rampton 2009). The goal in such quoted speech samples is not to give an accurate reproduction of the original tune (Bolinger 1946) but rather the size, slope and f0 range of the rise "can help to stylize quoted speech as marked speech" (Tyler 2014: 551). Thus, those learners that have learned to do that in the samples used in the present study could be described as more advanced learners who know how to use their f0 range appropriately to signal pragmatic functions and to stylize quoted speech.

Overall the learners show similar structural patterns in f0 range (RQ1), i.e. f0 level is native-like for HL-tunes and f0 span is narrower in HL-tunes for all learners. While the learners all show similar trends, the SpE speakers produce the narrowest f0 span in HL-tunes (RQ2). The learners come very close to native-like usage when it comes to f0 level (in HL-tunes) and entrainment effects (RQ3a) but deviate mostly in f0 span, which is narrower in non-native speech. However, even when measured in STs, f0 span deviates by SEX and TUNE_PATTERN (RQ3b). Thus, while learners produce a narrower f0 span in STs for the HL-tunes, they all produce a considerably higher f0 level and wider f0 span for LH- and HH-tunes (MurE speakers also produce wider spans for LL-tunes) than the NBrE intercept. The SpE speakers in general produced fewer LH- and HH-tunes than the CzE and GerE speakers, which explains why the SpE speakers in general have a narrower f0 span than the other learner groups. In sum, it depends which tune pattern is considered and which f0 range measure (span or level) is taken. Yet on the whole, the learners all exhibit striking similarities in their f0 range patterns. L1-specific and developmental issues will be discussed in the next section.

5.4.3 L1 influence or developmental factors in f0 range production

Czech, German, and Spanish have in common that they have been described to have a lower f0 range compared to English (e.g. Kelm 1995; Scharff-Rethfeldt et al. 2008; Estebas-Vilaplana 2014: 185; Volín et al. 2015: 111). An L1 transfer in f0 range studies is usually postulated when the interlanguage f0 range reaches intermediate values between the L1 and the TL (e.g. Mennen et al. 2014). However, some studies find even narrower f0 range in the L2 than in the respective L1 (e.g. Volín et al. 2015), thus finding evidence against L1 influence. Mennen et al. (2012) found German speakers to produce a compressed f0 range at the beginning of IPs and an expanded f0 range in later parts of IPs. This can be interpreted as L1 influence in the present study since the German learners were found to frequently use an L* H- pattern and even if the learners used low boundary tones, they were never produced with a low f0 as NSs would. Evidence for regional influence in the MurE f0 range has been shown to affect f0 range in that SpE intonation from Spain is exhibiting a quite narrow f0 range with flat intonation. According to the Atlas of Spanish intonation (Prieto and Roseano 2009-2013), the Spanish varieties spoken in Madrid and Murcia are part of the same dialectal region (Castellano centro-meridional), however, due to the geographical distance one can assume dialectal differences on prosodic realizations within these two regions. According to Monroy and Hernández-Campoy (2015: 237), the stress and rhythm of Murcian Spanish is the same as in Castilian Spanish, despite differences in intonation between the two varieties of Spanish. "In general, colloquial Murcian favours a narrow range of tones, to the extent that it may be considered one of the 'flattest' varieties among the Spanish accents, and one that may leave other native speakers, particularly South Americans, with an occasional impression of indecisiveness" Monroy and Hernández-Campoy (2015: 237). According to Monroy and Hernández-Campoy (2015), the level tones (together with narrow low-rises and low-falls) dominate in Murcian Spanish, while high-fall and high-rise tones are rarely used (237). Since the MurE speakers in this study showed consistent deviances from speakers from Madrid, either L1 influence of the regional variety can explain the deviances, or developmental issues may explain the variation, as will be discussed in the next paragraph.

Previous research found lower proficiency learners of English to produce a narrower f0 range, which, nevertheless, develops over time, i.e. it widens over time (Pickering and Levis 2002, as cited from Pickering and Litzenberg 2011). Furthermore, learners may learn to produce position-sensitive f0 range after having mastered the production of global f0 ranges (Mennen et al. 2014: 325). These studies seem to suggest that a narrow f0 range may diminish with increasing proficiency in an L2. The analysis of the study abroad variable can give further insights into prosodic fossilization, i.e. with prosodic deviances even after a stay abroad being evidence for features that are more susceptible to fossilization (see also Orrico et al. 2016). Since the present study is only pseudo-longitudinal, inferences on fossilization and developmental paths of individual learners cannot be made. However, as Orrico et al. (2016) suggest, those prosodic features that do not show any improvement in higher level learners can be thought of as more vulnerable to fossilization. F0 span, for instance, has been shown to be narrower in L2 speech, which can be explained by L1 transfer if the L1 exhibits a narrower f0 in general, and highly advanced learners of English have been shown to even 'overhit' the native f0 level target (Orrico et al. 2016: 128). This suggests that f0 range can improve over the years with increasing proficiency and is thus not likely to be vulnerable to fossilization. The prosodic cues that have been found to be possibly fossilized, because they deviate from native usage for all proficiency levels, were due to the realization of the f0 curve (rising, falling, flat, or combinations of these). All learner groups regardless of proficiency level have been found to use flat f0 contours the most (Orrico et al. 2016: 128) and realizations of the nuclear accent. Thus, while f0 range and prominence placement improve over time, the realizations of the global f0 curve and nuclear accents might still be susceptible to L1 transfer (cf. Orrico et al. 2016: 130). Another concept mentioned by Orrico et al. (2016: 129) is that of *prosodic drift*, "learners want to leave their L1, but they are not sure about what they should do because of the paucity of the input and the lack of direct instructions; thus, they drift away from the L2 norm in the attempt to reproduce something they believe is different from their L1". I have called this phenomenon "prosodic exploration phase", with the difference being that learners do try to reproduce native-like patterns but not in an aimless manner. They just try to be overly explicit to make their pragmatic intentions clear to their interlocutors. F0 span in the present study has been found to be most deviant in non-native speech (especially in MurE speakers) and a pseudo-longitudinal study has determined that f0 span is likely to fossilize, meaning that it did not improve after a stay abroad. For all tunes no improvement (narrower and lower f0 for LH- and HH-tunes and wider f0 span for HL-tunes) could be observed, suggesting that f0 span may also be fossilized, especially for GerE speakers who produce even wider f0 spans for LH-tunes after a short stay abroad. Therefore while for some learners these findings might also be interpreted as evidence for prosodic fossilization, i.e. those learners who have been abroad but still exhibit a deviant f0 span, for the SpE learners prosodic drift may be observed. Since the SpE speakers have been shown to produce the widest f0 spans in STs for HH- and LH-tunes (while the CzE and GerE have been shown to produce these tunes more often in general), this may be interpreted as an example of prosodic drift. This possibly means that the SpE speakers do not know how to produce the rises appropriately and may try to drift away from their native language as much as possible, which results in unnatural non-native-like production of high-register tunes. In L1 Peninsular

Spanish, f0 range was produced with a lower and narrower f0 range than English for some conditions, while f0 range may be higher and wider in other conditions (Sánchez-Alvarado 2020). The same results were found for certain tune patterns for the SpE speakers, which may point towards an example of negative L1 transfer from Spanish. As far as f0 level is concerned, there was no evidence that learners deviate from NSs when it comes to the HL-tune. However, they were shown to produce significantly higher levels for HH- and LH-tunes. Thus, the proficiency level and exposure to the TL are also factors that possibly influence the f0 range productions of NNSs and fossilization and prosodic drift may also be taking place (RQ3b). In sum, all learners show the same trends in f0 span by tune pattern, but the SpE speakers go into the most ‘extreme’ directions (i.e. narrowest f0 span in HL-tunes, and widest in HH-, LH-, and LL-tunes (MurE only)), while CzE and GerE speakers show more similar patterns with less extreme but still non-native-like values.

Furthermore, there is the problem of avoidance, where certain errors might be frequent in the speech of one learner and non-existent in the speech of another learner simply due to the avoidance of difficult structures in the L2. For instance, while one learner has a very flat f0 range and does not exhibit much intonational variation in his/her speech, he/she might not use intonation as much in general, while another speaker might make use of a wider f0 range and thus has the potential to make more ‘errors’ in the use of deviant f0 patterns for a specific sentence type. The complexity of an L2 category, along with a speaker’s ability to perceive it, might result in avoidance (Cruz-Ferreira 1983: 138). This point was already discussed in Section 5.2.3. To discriminate prosodic units and their meaning is not an easy task even for NSs (e.g. Scuffil 1982). Therefore, learners will have to be specifically trained to perceive different prosodic contrasts.

Pitch range of the 10 female strongly-accented CzE speakers in Skarnitzl and Rumlová (2019) also tended to exhibit a very flat f0 range. Studies such as Skarnitzl and Rumlová (2019) are very rare in which segmental and supra-segmental features are analyzed together. Together with some segmental deviances in the strongly-accented CzE speakers, f0 range is among the most deviant prosodic features (see p. 125 for an overview of all features). Their overview shows that, while almost all but one learner (1 out of 10) who speaks with a strongly Czech-accented English have an extremely flat f0 range compared to NSs. At the same time, individual speakers can perform native-like in some of the segmental and other suprasegmental features tested, while performing poorly in others. This implies that in order to assess a learner’s overall oral performance, one has to consider both segmental and prosodic features. Even at a strongly-accented level speakers struggle with different areas and perform well in some areas. There is considerable individual difference and combining the analysis of prosodic and segmental features can give further insights into L2 speech development. In the present study, however, fluency features did not play a significant role in determining f0 range differences in native and non-native speech.

The present study also found considerable individual difference in f0 range by all speakers, especially by the gender of the speaker. Even though the metadata of most existing corpora clearly divide the speakers into male and female, according to their f0 profiles, such a binary distinction is difficult to uphold. The unusual high f0 in male speakers and low voices in some female speakers pose a challenge in sociolinguistic prosodic variation research. Even in the speech of heterosexual adults, considerable interspeaker variation was found in the f0 (Weirich and Simpson 2018). The findings of previous studies suggest (van Bezooijen 1995;

Weirich and Simpson 2018) that f0 is also a learned component and not a biological one (Ohala 1983; Gussenhoven 2004) and even heterosexual men can have a more female f0 (Weirich and Simpson 2018). In sum, this seems to suggest that gender and individual speaker's differences seem to influence the f0 range either with or without suggesting a particular patterning based on L1 background.

5.4.4 Summary

This section discussed methodological and theoretical challenges, entrainment effects, as well as L1-specific phenomena, in order to analyze prosodic variation, when dealing with non-native speech as well as regional dialects of BrE and AmE, and gender-specific differences. The narrow f0 range of MurE speakers was explained by an interplay of regional and developmental aspects, and a higher proficiency has been suggested for the MadE speakers. Another point this section has made is that f0 range should not be used in isolation to assess L2 speech, because considerable variation in which features deviate from L2 norms has been demonstrated (see also Skarnitzl and Rumlová 2019). Additionally, f0 span was identified as the most deviant feature in non-native speech for HL-tunes, which does not seem to improve after a stay abroad and is, therefore, a fossilized feature. Learners, however, SpE speakers showed a tendency to use wider f0 spans in LH- and HH-tunes (and LL-tunes for MurE speakers), which were interpreted as instances of prosodic drift. While L1 influence cannot be ruled out as a factor that determines the production of narrower f0 spans for HL-tunes and wider f0 spans for high-register tunes, this could not be investigated with absolute certainty in the present study, especially since the learners from different L1 backgrounds all showed extremely similar trends in their f0 range patterns (level and span by tune), thus demonstrating similar developmental paths which are indicative of the learner's L2 proficiency levels. So far, the extremely high f0 range produced in LH- and HH-tunes in non-native speech have been mainly described as fulfilling a discourse management function (see Section 5.2.2) and as instances of prosodic drift (see Section 5.4.3) but other explanations for this tendency may apply, which will be discussed in the next section.

5.5 Discussion of uptalk results

The pragmatics of prosody will be explained in more detail by zooming in on the meaning, and use and non-use of uptalk in native and non-native speech. Predictions on uptalk were made with caution in light of the small body of L2 related uptalk research. In my opinion, the characteristics of non-native utterance-final rises have not been well understood and have not been dealt with in depth. Previous research attributed high boundary tones to uncertainty in non-native speech (e.g. Gut 2009) as well as in native speech (e.g. Nilsenová 2006; Shokeir 2008; Clopper and Smiljanic 2011; Levon 2016), while at the same time in native speech high edge tones were found to be used for a variety of pragmatic functions besides signaling uncertainty (cf. Barry 2008; Shokeir 2008; Ritchart and Arvaniti 2014; Armstrong et al. 2015; Arvaniti and Atkins 2016; Warren 2016; Levon 2016, 2018). In this study I have found uptalk to be used for a variety of pragmatic functions in non-native speech as well. It has yet to be established whether uptalk rises in native and non-native speech have similar forms and functions, because the NSs in this study did not use uptalk as much as the learners, or did not use uptalk at all. Therefore, this discussion is based on uptalk usage as it has been reported in previous research on SBrE and/or AmE.

The results presented in this section, however, do not mean that the other speakers of the corpus sample did not produce any instances of uptalk; it mostly means that the selected sample from the entire recording did not contain any uptalk instances or the speakers generally did not produce any uptalk instances. Furthermore, if one learner only had one instance of uptalk this does not mean that there are not any more in other parts of the recording (especially if only the part from Task 1 was selected). Mostly these uptalk tunes appeared within monologic parts and most frequently in checking and continuation, as well as a few instances of listing and to signal insecurity. While the frequency of occurrence might already give us insights about the use of uptalk in native and non-native speech, the frequencies might have been an artifact of the study design. Depending on the stretch of speech that was taken as a sample, uptalk might have been very frequent for a speaker and non-existent in another sample of the same speaker. While the total frequency counts of uptalk occurrences might have been a coincidence, a closer look at the different shapes of uptalk for the three different non-native English groups can give us insights into how these differ from each other and from native usage (based on results from previous research).

The following section (5.5.1) seeks to explain why the NSs of English, both AmE and NBrE speakers, did not use any instances of uptalk (or very few instances) compared to the NNSs in this study. Section 5.5.2 discusses why certain learners use uptalk frequently, and Section 5.5.3 seeks to explain whether L1 influence or developmental paths can explain the distribution and variation in the learner data.

5.5.1 Uptalk in native speech

In total, the NBrE speakers only used two uptalk instances (one female and one male). Nine out of the 45 AmE speakers (all female) use uptalk but only very infrequently. These include two speakers from the West of the USA, five SAmE speakers, and two New England speakers. Three out of these nine speakers are from the NWSP corpus and six from the NSV corpus. While much is known about AmE uptalking features in other speaking styles and contexts, not much is known about BrE uptalk. Compared to SBrE, SAmE, and AmE in general, there seems to be no research on Lancaster English prosody in particular (LOCNEC speakers are from Lancaster). The LOCNEC speakers clearly speak with an accent different from SBrE but also different from the other UNB varieties, e.g. Birmingham, Liverpool, Glasgow, Belfast, and Tyneside, who commonly produce rising intonation on declaratives (Cruttenden 1997: 133). For the accent spoken by the NBrE NSs in the present study, no such a trend has been observed. Previous studies on uptalk have shown that SBrE does not seem to pattern with other varieties of English such as AusE, AmE, and considerable differences are mentioned between a systemic and functional dimension of intonation (cf. Ladd 2008: 116; Arvaniti and Atkins 2016; Levon 2016, 2018). These differences are explained by a comparatively recent emergence of uptalk in the UK and that the feature has not yet grammaticalized (cf. Barry 2008; Levon 2018), as it has been reported in AusE for example. While some early studies suggest that uptalk has been used in the U.S. since the 1960s (e.g. Lakoff 1975; Ching 1982), Cruttenden (1994: 130) argues that uptalk did not arrive in London until at least two decades later. Levon (2018: 18) labels this the "[t]wenty-year lag", suggesting that uptalk has yet to develop a stable form and function in BrE. While uptalk has been found in SBrE, it seems to be completely absent from the speech samples of the Lancaster speakers in this study. The LOCNEC recordings are from 1995, only approximately 15 years after uptalk was first reported to arrive in London. If similar speakers would be recorded today, the variation might be different again and form-meaning mapping might be more adjusted for uptalk tunes. For instance, Lindsey (2019) notes that typical AmE features such as uptalk and vocal fry have become more frequent in modern BrE nowadays. Time wise, the other corpora compare to the NBrE as follows, i.e. 2000-2002 was the recording period for LINDSEI-SP, 2004 for LINDSEI-GE, 2012-2014 for LINDSEI-CZ, 1998-2004 for the NSV corpus, and 2003 for the NWSP corpus. While the time period might be one factor, I rather think that the non-use of uptalk mainly has to do with regional variation in the case of the NSs, where uptalk is not used. In sum, it is possible that the NSs do not use uptalk simply because it is not part of their repertoire, at least at the point in time of recording.

Besides regional variation, tones and tunes are very much influenced by stylistic variation and the context, as well as how distant the relationship to the interlocutor is (Nevalainen 1992). Previous research has shown that NE speech typically prefers HL-tunes or level tunes over rising ones (e.g. Grabe 2002). Nevertheless, in order to reduce social distance, rising tunes may be used more frequently (Nevalainen 1992; Tyler 2019 found rises to constitute only 6%, because his data includes intimate interactions to business transactions). Furthermore, in other discourse types, such as face-to-face conversations between equals or intimates, more falling tones are used, which is explained by the fact that no special appeal need be made to the listener (cf. Nevalainen 1992: 420). Some of the results in the present study may also be explained by the *Bulge Theory* (Wolfson 1986a: 694, b), which explains interactional behavior between interlocutors of different gender-pairings. According to the Bulge Theory, greater needs of negotiation occur when interlocutors are vaguely familiar with each other and that they possibly may become friends or more familiar with each other. In my study, I think the relationships between the interviewers and interviewees should be fairly clear. Unfortunately, the interpretation of these findings is complicated by a lack of information about the level of familiarity between the NSs. It can be deduced by the context or the form of address that they differ in their degree of familiarity. In the SAmE corpus, the interviewers are status-equal and may even be friends, which may explain the more informal style at times. However, also highly formal interviews with a strict protocol of questions can also be observed in the present study, such as the NWSP recordings for instance. Uptalk is only used when a hearer is present (Lindsey 2019: 108), maybe the AmE speakers in the monologues are not really interacting with an interviewer due to the task setting or speaking style. The frequency of uptalk in NBrE and AmE could be connected to the type of data the corpus of the present study consists of. Some speaking styles are more likely to elicit uptalk instances. For instance in map tasks (e.g. Barry 2008) uptalk has been reported to be more frequent than in dialogues or free speech. A map task requires constant monitoring and information requesting (cf. Barry 2008: 217). It would be interesting to see whether the NSs from Lancaster, if recorded today and if involved in a map task, would produce instances of uptalk.

Another interpretation why the NSs do not use these extreme high boundary tones on their declaratives, as found in the NNS samples, is because of the way the speakers possibly perceived the interview task. Previous research found that in self-conscious speech or speech with increasing levels of formality, NSs of English tend to produce speech which is perceived to be the standard language (i.e. the RP prestige variety) and thus use less regionally marked forms in more formal contexts (e.g. Lowry 2002; Ulbrich 2008; Tyler 2019). In most English varieties, uptalk is still a stigmatized linguistic feature attributed to young women expressing uncertainty. Even in Belfast English, where rises are predominantly produced on declaratives, for most speakers, the more formal the situation, the more falls are produced and rises are more frequent in informal styles (cf. Ulbrich 2008). The more informal an interview is, the larger the rises at the end of utterances are (Tyler 2019). Thereby, Lowry (2002: 33) suggests that female speakers seem to be more willing than males to use the features of the prestige variety. Male speakers were found not to use uptalk as often (see also Britain 1998: 226), where male participants in an interview were found to be less responsive to the interviewer and might have perceived the interview as a task to be completed as fast and efficiently as possible and, therefore, do not engage in genuine and prolonged conversation. Since uptalk has been shown to be used in longer stretches of speech as opposed to the shorter stretches of speech by many male participants, this could explain their absence of uptalk. In the NE recordings, there are more male speakers as compared to the non-native corpora selected for the present study. Citing claims from Ochs (1992), Tyler (2019) argues that the reason why female speakers produce more and larger rises could be explained by the fact that women interpret interactions in more social terms. However, it is possible that the perceived degree of formality of the task by the different speakers might overall reduce the prosodic variation, especially regional prosodic variation in the NSs of English in this study, who either did not (yet) adopt uptalk usage or simply suppressed it in the interviews. The NSs may be more aware of the stigma uptalk accrues and might be motivated to avoid it. In previous studies on segmental features conducted by Labov (1972), style and education have shown to affect the production of regional variants. For a discussion of the degree of formality in the present data set, see Section 3.8. Thus, in sum, another explanation might be that the NSs and learners might have perceived the interviews differently, and the NSs are more aware of stylistic variation of regional prosodic features and the stigma that uptalk carries. There might also be gendered-differences in the perception and execution of the interview task.

It has been speculated that uptalk originated because of the negotiation of meaning with and between NNSs in immigration settings (cf. Guy and Vonwiller 1989: 33). While it has been commonly found that native and non-native speakers use rises as meaningful prosodic cues in interactions to negotiate meaning and to ask for feedback (e.g. Pitzl 2005), there are hardly any such uses of rises in the LOCNEC samples. The interviewer in all LOCNEC files is female and non-native with a native-like proficiency in English. If they had talked to a NS fellow student and/or a friend, the outcome might have been different. However, even though this is the case in the AmE recordings, where the NSs are interviewed by fellow students and generally the atmosphere seems to be more relaxed, uptalk is not used in the same way as in the non-native samples. The NSs do use high boundary tones but they do not necessarily seem to elicit feedback or approval from the interviewers, as is the case with the NNSs. Also, if the NBrE speakers use uptalk at all in their speech besides the interview, the question is whether they would use it with a non-native interviewer at a university setting. Since the uptalk feature was a relatively new arrival in England at the time of recording and not yet grammaticalized as in AmE or AusE, the NBrE speakers might not have wanted to confuse the interviewer by using uptalk, which might be mistaken to signal questions (see example of a misunderstanding caused by uptalk in Section 5.3.1). Generally, the interviewer threatens the interviewee's turn-at-talk in every single interview by intervening and asking questions at every signal of pausing. Possibly this is another reason why the NSs did not need to ask for approval or comprehension checks because the interviewer constantly backchanneled, overlapped with their speech, and interrupted them to take over a lengthy turn. The interviewer's overexcitement might have inhibited the use of rises to signal continuation and comprehension checks were seemingly not needed because of the over-attentive nature of the interviewer. In sum, the interviewer's status, familiarity, personality, and interview style could possibly also have an influence on the use and non-use of tunes.

Thus, if the NSs do not use uptalk in their interactions to negotiation meaning or to bond with the interviewer, the question is what other linguistic features do they use instead? As will be shown in the discussion on pragmatic prosody in Section 5.5.2, pragmatic markers or discourse markers are linguistic devices that can also serve the purpose of negotiation of meaning and bonding. NSs have been shown to use pragmatic markers such as *you know* and *I mean* more often in LOCNEC (Buysse 2017) and the LLC (Romero-Trillo 2014) than NNSs, this might point to different preferences in how the bonding or reduction of social distance are realized in interactions. In addition to the connection between pragmatic markers and intonation, connections between uptalk and tag questions have been made by several researchers (e.g. Allan 1990; Holmes 1995; Tench 2003; Ramírez Verdugo and Romero-Trillo 2005). Further dialectal differences have been described in the case of tags and uptalk in AmE and BrE. Tottie and Hoffmann (2006), for instance, found in a corpus analysis that tags are nine times more common in BrE than in AmE. Based on previous research, they conclude that AmE may make more use of intonation instead of using tag questions. Some researchers suggest that what may have once been a tag in its lexical form has remained only as the tune in AmE, i.e. a type of linguistic abbreviation (Tench 2003). Thus, while both NS groups in the present study may make greater use of lexico-grammatical features as pragmatic markers and tag questions to connect with an interlocutor, AmE speakers have been described to make greater use of intonation, which was corroborated in the present study on the prosodic level.

This section has reviewed reasons for the low frequency or non-use of uptalk in general in the NS samples. It can be concluded that the feature might not (yet) be part of the NSs' linguistic repertoire, if it is in their linguistic repertoire, that they might suppress it in more formal speaking styles, that its usage depends on the context, perception of the interview task, and the interviewing style, and the relationship to the interviewer, and status differences. Additionally, NSs might use other linguistic devices such as pragmatic markers to signal similar functions instead.

5.5.2 Uptalk in non-native speech

The present section seeks to answer the following questions in the respective four paragraphs, i.e. 1) *Is uncertainty the full picture or is uptalk used as a strategy to gain mutual understanding and to prevent misunderstanding?*, 2) *Why would learners be more uncertain? and can gender explain the variation?*, 3) *What is the relationship between uptalk and* *creating coherence in utterances?*, and 4) *Are learners using a wrong intonation and creating a wrong pragmatic intent?*

Uptalk in native and non-native speech has been frequently attributed to uncertainty and nervousness (Gut 2009; Hoff 2014: 58). As argued by previous researchers for uptalk in native speech, which has and still is a stigmatized linguistic feature, it can be acknowledged that "[u]ptalk may indeed indicate uncertainty, but it simultaneously signals much more" (Linneman 2013: 83). While uptalk researchers have provided evidence that uptalk is not only used to signal uncertainty, this has not been shown for NNSs so far. One of the few studies that investigated uptalk in non-native speech (Flaig and Zerbian 2016) found that learners not only used uptalk to express uncertainty (usually about a vocabulary item) but also for floor-holding. The use of high boundary tones at the end of declaratives was used for various pragmatic functions in the present study, i.e. mainly continuation and checking. One of the main arguments the present study wants to make is that uptalk in non-native speech does not only signal uncertainty but an array of different pragmatic functions, in addition to discourse markers and connectives to create coherent units and to involve interlocutors in narratives and interactions. This indicates that learners in the present study are capable of using linguistic features to signal different pragmatic functions through the use of various linguistic features, including prosody. Therefore, a simplification of high end of utterance intonation in non-native speech to signal uncertainty or nervousness cannot give a complete picture. The use of uptalk may be a signal of appropriate pragmatic use but also of deviant usage. Whether the relatively higher usage of uptalk by the learners can be explained by uncertainty alone will be investigated in the present section. The present section specifically addresses the following questions: *What might be the reason for learners being more uncertain than NSs? And what alternative interpretations are there?* Several linguists have argued that uptalk has to be interpreted relative to the context in which it occurs (Gunlogson 2001; situational and temporal context: Tomlinson and Fox Tree 2009). The following utterances (examples (8)-(11)) made by CZ032, when talking about an experience in her monologue, illustrates the different uses of uptalk (underlined words with question mark) in an individual speaker:

(8) CZ032: "and when I became old enough I started to go as a <u>councellor?</u>"
(9) CZ032: "and actually um I would say it's sort of like a <u>writing? experience for me?</u> because since I was sixteen I made almost all of those games except for like two"
(10) CZ032: "so I would say it really gives me opportunity to I would say <u>explore myself?</u> from on=on different levels"
(11) CZ032: "and um it was fun because I wa=needed to learn how to be <u>diplomatic?</u> and I knew the councellors that were before me so I was really conscious about their <u>mistakes?</u> and I was trying not to make them"

Example (8) demonstrates that uptalk is used as something like new information. No backchanneling occurs after the uptalking tune. In example (9) a sustained rising tone on "experience for me" can be heard. In that case uptalk is used as a confirmation request to integrate the interviewer in the conversation and to establish common ground, something along the lines "do you know what I mean?" and it is acknowledged by the interviewer with a "right". Example (10) is similar in sound and pragmatic function as in example (9), which is also followed by backchannel *mmh*. Here the uptalk tune has more of a hint of "can one say that?" maybe this one can be more uncertainty of whether one can say that but also to seek confirmation of comprehension. In example (11) both uptalk rises are followed directly by backchanneling by the male interviewer. Here the uptalk instances have a hint of uncertainty about the words used but at the same time they fulfill a checking function. These examples from the monologic part show that the speaker uses rises at the end of IUs to bond with the interviewer, to include him, to check for comprehension, and to continue holding the floor. These examples are no clear signs of uncertainty, because the speaker is holding this information. Also, the speaker does not show any signs of difficulties with lexical retrieval of words in her speech, and has been rated to have a CEFR level of C1, indicating the end of utterance rises cannot be interpreted as signals of uncertainty only. In addition to using deviant prosodic patterns with pragmatic functions, learners may prefer using other resources to signal uncertainty, i.e. lexico-grammatical features such as modal verbs and adverbs (Ramírez Verdugo 2005). Interestingly, according to Buysse (2017: 40), the pragmatic marker *you know* functions (among other functions, cf. page 42) as a social monitor for conversation management (floor holding and yielding) and comprehension checks. This resembles the functions fulfilled by uptalk in the present study. However, learners in LINDSEI, according to Buysse's (2017: 52) analysis, use *you know* for a comprehension-securing function only very rarely and it is never used by the NSs, possibly because they use intonation to fulfill that function (my interpretation). This confirms my findings in that the NBrE (LOCNEC) NSs also rarely or never use intonation for such a function. Previous research has found that learners vary their use of discourse markers (e.g. *well*, *like*) and smallwords (e.g. *sort of*, *kind of*) less, compared to NSs (Götz 2013). NBrE speakers (LOCNEC) use the pragmatic marker *you know* almost twice as much as SpE and the GerE speakers use it very infrequently (smaller than one fifth of the native control corpus) (Müller 2005; Buysse 2017), and learners use the pragmatic marker *you know* with non-native-like intonation, i.e. with avoiding rising tones with pragmatic markers (Romero-Trillo 2014). One explanation could be that the learners use rises elsewhere, not realized on the pragmatic markers and maybe they deem the use of the marker itself as sufficient and, therefore, they do not realize that intonation is also used in combination with the pragmatic marker as NSs would. Thus, while German learners may possibly use more uptalk to compensate for the less varied use of discourse markers and fewer instances of pragmatic markers (*you know*), the Spanish learners seem to rely more on native-like patterns and consequently their use of uptalk is reduced. Interestingly, Romero-Trillo (2014: 214) found the female speakers from LINDSEI-SP compared to the female speakers from the LLC to use *you know* less but *I mean* more, according to him this is an indication of the "need for self-explanation" in the NNS-NS interactions. Not only did the learners deviate from the frequency of the two pragmatic markers, but also in the prosodic patterns they are realized with. According to Romero-Trillo (2014: 216), NSs use pragmatic punting for politeness reasons only in case a repair is needed, but they are generally certain that an interlocutor can follow what is being said. NNSs, on the other hand, feel insecure about their ability to convey their thoughts and, therefore, make different pragmatic and tonal choices to indicate that the message may have to be clarified or reformulated to prevent misunderstandings. Future research may have to determine what other resources learners use to signal uncertainty besides intonation. The analysis of lexico-grammatical features in combination with gestures and gaze should yield interesting results. How learners prosodically modify their interlanguage while interacting in L2 discourse with native and other non-native speakers has been analyzed in previous research mainly from a structural perspective. The functional perspective of prosodic modifications to L2 speech and the negotiation of meaning have not been analyzed frequently. Negotiation for meaning (NfM) means a lot of things to different researchers (see a summary of the different views in Foster and Snyder Ohta 2005: 405). While for some researchers the negotiation for meaning is something to repair communication breakdowns, i.e. something negative (e.g. Pica 1994), others see it as a positive co-constructional feature that aids communication (e.g. Foster and Snyder Ohta 2005). Previous research on the negotiation of meaning in NS-NNS and NNS-NNS interactions has focused on a variety of features except for intonation (e.g. Long 1983; Gass and Varonis 1985; Varonis and Gass 1985; Pica 1987; Pica et al. 1989; Pica et al. 1991; Foster and Snyder Ohta 2005; Pitzl 2005). Although some of these studies note that intonation also plays a role in the negotiation of meaning (i.e. rising intonation to e.g. clarify or confirm the meaning of a message), it is just mentioned in passing, and only few studies specifically look at the form and function of intonation in the negotiation of meaning (e.g. Pickering 2009; Pickering and Litzenberg 2011). Brazil's (1997) discourse-pragmatic approach to intonation supports the view that interlocutors use intonation to negotiate a "common ground" of understanding. Within Brazil's (1997) model of intonation, Pickering (2009) argues that tonal choice and relative pitch level contribute to intelligibility and interactional success in EFL interactions. Pickering (2009) and Pickering and Litzenberg (2011) found that EFL interlocutors orient towards tonal choice and f0 range to interpret and negotiate meaning. Both features are used to signal trouble spots and to negotiate their resolution. Just as in Foster and Snyder Ohta's (2005) study, the learners in the present data set also do not have any communication breakdowns, rather "they are sharing their meanings while monitoring and modifying their own and each other's utterances, minimizing overt communication breakdowns, and the accompanying frustration" (Foster and Snyder Ohta's 2005: 425). Miscommunication happens if a misunderstanding is not resolved after several turns (Dascal 1999: 754). In the CzE, GerE, NBrE, and AmE conversations in the present data set, clarification requests are rarely uttered by interviewers or interviewees. Sometimes, non-understanding is even ignored in order to not disrupt the conversational flow, e.g. in personal communication with one of the interviewers from LINDSEI-GE, one can hear that one learner (GE039) misinterprets the word "climate" for "climax" and goes on to talk about a different topic (an encounter with a shark and other highlights of his trip, i.e. a volcanic eruption) than the interviewer intended, i.e. to talk about the weather in the country. The interviewee goes on for some time and no efforts are made by the interviewer to correct the misunderstanding. This is indicative of a face-saving-strategy employed by the interviewer (elicited from personal communication with the interviewer).

A: What was the climate like?
B: the climate was the climate was meeting a shark which was about my size
B: […]
B: um another impressive climate of my journey was um seeing a volcano eruption

The interviewer told me in personal communication that she did not want to interrupt the student even though he went on to talk about a completely different topic. In addition, the learners are quite advanced and there are rarely opportunities that require clarification requests. Therefore, my results do not compare to those with frequent clarification requests in interactions such as Pica et al. (1991), because these were based on low-intermediate level students in information gap tasks where clarification and modification of utterances are important for successful task completion. The task in the recordings in the present study was to "get the students" to talk as much as possible. Therefore, misunderstandings are rather rare but one can rather find several attempts at gaining mutual understanding and preventing misunderstanding (cf. "proactive work in talk": Mauranen 2006: 135), i.e. by the use of uptalk. Learners in the present study made frequent use of high edge tones with the function of confirmation checks (called "checking" in this study), which have been found to be used for the prevention of misunderstanding (Mauranen 2006: 136). Additionally, learners were also found to use uptalk to check whether the listener is able to follow longer narratives in monologic speech (see also Guy et al. 1986). In sum, uncertainty is not the only explanation for the occurrence of uptalk in non-native speech and the prosodic feature is indeed used for a variety of pragmatic functions to fulfill certain discoursal and social functions. Furthermore, the frequency of pragmatic markers and the prosody they are realized with can give further indication of a mismatch with native-like speech. Thus, while learners may have the same intentions (functional level) in conversations like NSs, they deviate on the distributional and realizational level on different linguistic levels, i.e. lexico-grammatical features and prosodic patterns. This deviance can be explained by a higher need on the side of the learners to make themselves understood, while NSs take for granted that they are being understood. Therefore, while uptalk fulfills various pragmatic functions, uncertainty about being understood and to prevent misunderstandings is certainly also part of the pragmatics of uptalk in non-native speech.

The next question to be addressed is why learners (especially female learners) would be more uncertain than NSs. Of course, learners do have a good reason to be more uncertain about their L2 than NSs considering they have to correctly retrieve lexical items, form their utterances spontaneously, and if they can manage that, they may also pay attention to their correct pronunciation and intonation. Gender is another factor that may explain the results on uptalk usage. The use of uptalk in female native speech is also generally perceived as signaling uncertainty, which may index general gender-stereotyping. Previous research on native uptalk usage has shown that while both sexes use uptalk, they may use it with a different distribution, i.e. female speakers use uptalk more for floorholds, because they are generally interrupted more (Coates 2004; Ritchart and Arvaniti 2014). However, in the present study there are not many male interviewers. In LINDSEI-SP there are only female interviewers and in LIDSEI-GE there is only one male interviewer (an AusE speaker who uses uptalk). Only the CzE corpus is balanced in terms of the gender of the interviewer. This is why the effect of gender-pairing is not analyzable with the current data (at least for any reliable results across the different groups). Cross-varietal studies of English have found that women produce uptalk rises differently than men, i.e. they produce them more frequently, they start their rises later, they produce them with a larger pitch dynamism, and that they use them for different pragmatic functions (e.g. Ritchart and Arvaniti 2014; Armstrong et al. 2015; Arvaniti and Atkins 2016). Thus, there seem to be gender-based differences in native speech on different prosodic levels, i.e. functional, distributional, and realizational. According to Ochs (1992: 340), "few features of language directly and exclusively index gender" and the "relation between language and gender is distributional and probabilistic". This was also found in this study; none of the prosodic features analyzed was exclusive to one of the sexes (see also Arvaniti and Atkins 2016). Instead, rather female speakers produced uptalk features more frequently, and also there were proportionally more female learners than male ones in the present study. According to Smith and Clark (1993: 26), answering questions has a social function, i.e. the exchange of information and self-representation. They relate their postulation to Grice's (1975) conversational maxim of quality, according to which speakers stick to the truth content of their utterances. Thus, the authors contend that if a speaker expresses uncertainty in an answer prosodically (and visually), he/she does that in order to preserve the maxim of quality, in Gricean terms. Maybe the female learners are more communicative in general and picked up the uptalk feature from NSs, and it evolved out of communicative use. Research by O'Barr and his associates (Conley et al. 1978; O'Barr and Atkins 1998 [1980]) found that non-expert and powerless speech (lower social status) in courtroom witnesses was characterized by a frequent use of hedges, hesitations, intensifiers, polite forms and question intonation "making a declarative statement with rising intonation so as to convey uncertainty" (Conley et al. 1978: 1380), which is associated with female speech in general (cf. Lakoff 1975). This overrides gender-differences, in that neither female nor male witnesses used such features if they were experts on a topic and/or more powerful in their status. Interestingly, Levon and Ye (2020) found that men are perceived as more likeable and trustworthy when using uptalk as defendants in mock rape trials, since they use uptalk to ask for inclusion in the interaction. These findings might extend to other contexts outside of the courtroom as well, where speakers are more certain on their topic and this could be reflected in their voice and use of intonation. The learners who use uptalk frequently in the present study are experts on their topics, because they include information that only they themselves hold and the interviewer cannot confirm or try to verify the truth content of the statements in any way. Interesting findings have also been found in Linneman (2013: 96), who investigated the use of uptalk in the Jeopardy! show. He found that the more successful women were in the show, the more uptalk they use to reproduce gender order. Other research has shown that dominant women are considered rather unlikable and unfeminine (Costrich et al. 1975; Carli 1990; Heilman and Okimoto 2007; Williams and Tiedens 2016). Linneman (2013) concludes that the women in the show use uptalk to create the appearance of uncertainty of their knowledge and to apologize for their success, while men rather want to appear competitive toward other men by using less uptalk. Maybe the learners in the present study are aware of the stereotypes surrounding women, dominance, rudeness, and the perception of the German accent coming across as rather aggressive and they try to mitigate the negative effect by using more rising tones and thereby sounding more likeable. Furthermore, I speculate that the phonetic differences in the uptalk rises produced by the learners may contribute to the perception of uncertainty rises, i.e. possibly because of the production of steeper and later rises. Further perception experiments would be necessary to test how listeners perceive uptalk to be different in native and non-native speech.

An alternative explanation to why learners use uptalk more frequently than NSs in the present data is that uptalk is possibly used to create coherent units, to plan utterances ahead and buy more time to formulate answers (see discussion on intonation and fluency in Section 5.3) (Foster and Snyder Ohta 2005: 413; Tomlinson and Fox Tree 2011: 61). While NNSs of the LINDSEI-SP, and -GE corpus have been shown to use lexical pragmatic markers[78] such as *you know* less frequently than NSs (LOCNEC), they were still found to use it for all of the proposed functions of the pragmatic marker as in native speech (cf. Buysse 2017; see also Romero-Trillo 2014 for the same results for LINDSEI-SP compared to the LLC). Interestingly, a study by Buysse (2017: 46) found that Dutch, French, German, and Spanish learners of English used the discourse marker *you know* with additional connectives to create a coherent unit with the next utterance more often than NSs. The Spanish learners did that with 70.37% of the discourse markers, the Germans with 52%, and the NSs with 35.9%. The author explains this by a possibly "stronger desire among the learners to make coherence relations explicit" (Buysse 2017: 46). Neary-Sundquist (2008) found English learners of German speakers to produce more conjunctions instead of discourse markers when involved in a narrative task. Additionally, rising f0 has been found to signal cohesion between prosodic units (Selting 2010: 8). One characteristic of uptalk is that it occurs in a series that a "succession of HRTs" functions as a cohesive device to maintain "the integrity of the content and the continuity of the speaker turn" (Britain and Newman 1992: 10; Wichmann 2000: 93, 85-86 calls this phenomenon "intonational parallelism"; Crystal 1969: 241 calls it "tonal reduplication"; Bolinger 1989: 205 "series intonation"; see also Linneman 2013). A series of uptalk tunes may occur in adjacent as well as non-adjacent IUs, but was found to be much more common in consecutive IUs (Wichmann 2000: 87-90). The same tendency was observed for the learners in the present study, if uptalk occurred, it was very likely to occur in a series and in consecutive IUs. Thus, learners use *you know* with connectives more often but they also use it in connection with uptalk to make coherence relations explicit, while NSs might deem a pragmatic marker such as *you know* as sufficient for such a purpose. Uptalk tokens in Arvaniti and Atkins (2016) were mostly utterance-final and only 3% of all uptalk instances were ip-final. The present study, however, found NNSs (men and women) to use uptalk over 55% IP-finally and over 44% ip-finally. An explanation might be that NSs only feel the need to signal cohesion at larger constituent boundaries, while NNSs might need to work in smaller steps. To give one example how prosody and fluency features can complement each other, of all uptalk instances analyzed in the present study, 50% in female and 55.17% in male speech were followed by a breathing pause, 28% in female and 27.59% were followed by silent pauses, and 21.92% vs. 17.24% were followed by no pause. Since most uptalk instances in both male and female speech are followed by breathing and no pauses, this might indicate that the learners are not asking questions or really asking for feedback but rather use it as a cohesive device.[79] Interestingly, 96.92% of all female uptalk instances and 100% of the male instances are not interrupted by an interviewer question or comment. This shows that rarely any of the interviewers perceived these rises as questions or an invitation to make comments. Whether the uptalk features were followed by backchanneling or non-existent or non-audible backchanneling was 50/50 in both male and female interactions with interviewers. This might explain why learners continue using uptalk in the following utterances, i.e. to give the interviewer another opportunity to ask questions or to provide backchanneling at these points (see also Sacks and Schegloff 1979: 19-20). This example shows that fluency features should

[78] See Romero-Trillo and Newell (2012:120-122) for a discussion of the difference between pragmatic markers and discourse markers.

[79] See Wichmann (2000) Chapter 4 for an overview of intonation as a cohesive device.

be combined when analyzing uptalk in dialogues because they can give additional insights into the behavior or perception of non-native speech by interlocutors, because it is not only the intonation that listeners pay attention to, but also an ensemble of segments, fluency features, lexico-grammatical, and prosodic features. Thus, to conclude this train of thought, learners use uptalk for purposes other than signaling uncertainty, i.e. to signal cohesion, just as much as they may indeed express uncertainty with the same rises. This can only be achieved if pragmatic functions of tunes are included in a study. A simple list of tunes alone cannot give a full account of the situational usage of these tunes. Determining the pragmatic function, however, is a relatively difficult task as the next paragraph explains.

Another problem is that learners might produce intonation and intend to use other intonation and they are not even aware of their use of intonation to signal a certain pragmatic function. Pike (1945: 22) claims that "we often react more violently to the intonational meanings than to the lexical ones; if a man's tone of voice belies his words, we immediately assume that the intonation more faithfully reflects his true linguistic intentions". Therefore, learners might be misunderstood as to their linguistic intentions if their prosody does not match the pragmatic function. It has been demonstrated that NNSs differ in their tonal choice to indicate degree of probability. For instance, Spanish NNSs have been found to produce a falling tune even in contexts where they are signaling uncertainty and rising tunes in contexts where they were not uncertain. Thus, a mismatch between lexical signaling of uncertainty and the use of intonation were found (Ramírez Verdugo 2005). Previous research has found that even advanced learners of English still have problems with prosody and the pragmatic meanings it conveys in their L2, and overgeneralizations of falling tones to signal different pragmatic meanings is very frequently reported (e.g. Toivanen 2003; Ramírez Verdugo and Romero-Trillo 2005; Ramírez Verdugo 2005; Ramírez Verdugo 2008; Pickering et al. 2012), in addition to an overgeneralization of rising tones in tag questions and level tones with pragmatic markers, which NSs of English vary intonationally to signal different nuances in meaning (Ramírez Verdugo and Romero-Trillo 2005; Puga et al. 2017, 2018). In the non-native speech samples, mostly level and a few fall tonal patterns are produced (Romero-Trillo 2014: 218). Thus, the learners might have signaled finality or certainty but their wrong tonal choices may lead to the perception of uncertainty. Thus, to conclude, learners may not be aware of the pragmatic functions of rises and falls and may overgeneralize them to inappropriate contexts of use and may thus appear more uncertain even though they may not intend to do so. The true pragmatic intent by the learners in the present study is, however, not testable.

5.5.3 L1 influence or developmental factors in uptalk usage

Interestingly, no larger phonetic differences in f0 range of uptalk tunes were found in the CzE and GerE utterances, expect for GerE speakers with no stay abroad tending to produce wider f0 spans and CzE speakers with long stays abroad tending to produce steeper relative slopes of the uptalk rises. Also, the notion that late peak uptalk tunes (L* L-H%) and earlier peak uptalk tunes (L* H-H%) are consistently used with distinct pragmatic functions in non-native speech could not be supported in the present study. These findings suggest that learners may follow a common developmental path in the acquisition of prosody. The present section summarizes and contrasts the findings of the present study with the findings of linguists who mainly explain these high frequencies of high edge tones by either uncertainty, a lower proficiency level, and/or by direct L1 transfer. Götz's (2013) perception study showcases the general opinion on high boundary tones in L2 speech by native listeners. Götz (2013) reports on NS ratings (n=50) of 5 learners from LINDSEI-GE, which she chose on the basis of different parameters. One of these learners (GE041) was selected according to her strong German accent and intonation. The raters rated these five learners on a 10-point scale from 1 (sounds like an absolute beginner) to 10 (sounds like a NS). Very frequently, intonation and accent were one of the factors that led the raters to give lower fluency ratings. Particularly interesting were the comments raters made on the use of HRTs by learner GE028 who was chosen by Götz on the basis of her medium number of errors and fluency performance. One of the raters, for instance, commented: "what marks this speaker out as a NNS is her intonation which is full of "high rising terminals" but in some way distinct from the way they appear in the speech of e.g. young Australians or New Zealanders. Not sure what the difference is, but she sounds very foreign", while another rater comments: "Her end of sentence intonation is off most of the time. It gives her away as a German" (comments taken from Götz 2013: 154). These comments show that the raters in the perception study noticed the HRTs, but think that they are different from the ones used in for instance in AusE and give them away as German. As mentioned in the previous section, high boundary tones in L2 speech were mostly associated with uncertainty by previous researchers. As discussed in Section 2.8.2, uptalk is virtually non-existent in the L1s of the learner groups analyzed in the present study and can therefore not be ascribed to an L1 influence. However, it is still a valid question to ask *how uncertainty is prosodically expressed in L1 Czech, German, and Spanish and whether the uptalk examples can be, therefore, attributed to an L1 influence?* An attempt to answer the second question is made in the following paragraph.

Unfortunately, very little is known on how uncertainty is intonationally expressed in native speech (Ward and Hirschberg 1985; Wollermann 2012) as well as non-native speech (e.g. Ramírez Verdugo 2005). High f0 can be associated with friendliness and politeness on the affective level of the *frequency code* (Gussenhoven 2002), while a low f0 may be associated with aggression. On an informational level of the frequency and production code, high f0 is associated with uncertainty and continuity and low f0 with certainty and finality (cf. Gussenhoven 2002). Generally, in all three L1s of the learners in the present study, high tones are used to signal uncertainty. However, rising intonation is not the only way to signal uncertainty in these L1s. From previous research we know that the lower the level of certainty, the more hedges, hesitations, filled pauses, delays, slower speaking rate, and rising intonation is used (cf. Smith and Clark 1993; Swerts and Krahmer 2005; Wollermann 2012, with read German speech). Also, these linguistic means are not the only channel through which speakers convey their level of certainty, but they may also use facial expressions and eyebrow movements to support their message (cf. Cavé et al. 1996; Wollermann 2012). According to Wollermann (2012: 30, 100), uncertainty has an expressive component, which is produced and perceived through various modalities, i.e. fluency, facial expressions, gestures, and prosody and the non-verbal signals can support, replace or contradict the verbal signal. Speakers as well as listeners use both the verbal and non-verbal channels to make sense of the utterances and their content. While facial cues and eyebrow movements contribute to the perceived prominence in utterances, prosody (i.e. pitch accents) has been found to be the most dominant cue (e.g. Swerts and Krahmer 2004, 2005; Scarborough et al. 2009). However, auditory and visual cues contribute together to uncertainty perception, with visual cues being stronger than auditory effects (cf. Swerts and Krahmer 2005; Dijkstra et al. 2006). How uncertainty rises are different from other types of rises is not very well-known. Ward and Hirschberg (1985) contend that it is important to distinguish fall rises to signal uncertainty and other fall-rise tunes. Ward and Hirschberg (1985) maintain that the fall-rise tune in AmE to signal uncertainty is scooped, has a late peak in an accented syllable with an abrupt drop in pitch on the following two syllables, and the rise in pitch occurs sentence-finally (784). According to Ward and Hirschberg (1985: 765), "A speaker need not actually be 'uncertain' in order to use FR [fall-rise]: it may be used to convey uncertainty for purposes of politeness, irony, or deference". Generally, uncertainty is not expressed on a single word but rather through the prosody of the preceding context and other signals such as facial expressions (Liscombe et al. 2005; Pon-Barry 2008; Pon-Barry and Shieber 2009; as summarized in Swerts and Hirschberg 2010). In sum, in German, and English native speech uncertainty is said to be produced with rising intonation, although this was only analyzed in unnaturally produced read speech (cf. Pon-Barry 2008; Litman et al. 2009; Wollermann 2012).

Another insightful area may be the intersection of prosody and narrative structure and the social construction of interactions. Jucker's (2008) found in his study on movie retellings of German students from Giessen university, speaking either German or English and AmE speakers from California State University that the German learners of English adopt pragmatic strategies (i.e. sequencing of the narrative, choice of tense, introduction of characters, and reported speech and thought) from their L1. Jucker (2008) found that there are differences in how movie narratives are realized by AmE, German, and GerE speakers. He found that the advanced GerE speakers tend to adopt pragmatic strategies (i.e. structuring of narrative elements and their choice of tense, introducing characters and reporting of speech and thought acts) from their L1. For instance, while NSs of English mostly stick to one tense within a narrative element, NSs of German and German NNSs of English tend to switch tenses within such elements or they stick to the same tense throughout the entire narrative. These findings could also extend to prosodic features and pragmatic strategies in the present study. For instance, Warren and Britain (2000: 161) report on a study of the placement of HRTs within narratives, where they found gender differences with women using HRTs mostly in the orientation and evaluation portions of the narratives, men used them mostly in the complicating actions. The authors explain the difference by referring to the function of HRTs to establish solidarity and reduce social distance, which women make use of at the beginning of their narratives (see Levon 2016, 2018 for a similar discussion of the pragmatic functions of HRTs in London English). This was a feature often found in the non-native samples in the present study as well, where some speakers only use uptalk at the beginning of their narratives or speaking proportions. Thus, while an L1 influence could not be investigated in the present study in narrative structure, one can, however, observe a gender-based difference which patterns much like NE in narratives. The learners may use uptalk at the outset of their narratives to reduce social distance and to come across as more polite in order to "facilitate social interactions" (Astruc and del Mar Vanrell 2016: 2).[80] According to Britain (1998: 215), narratives are more "other-oriented" because they often include the sharing of personal

[80] Studies at the interface of politeness and prosody as used by L2 learners are extremely rare (e.g. Astruc and del Mar Vanrell 2016). See Brown and Levinson (1987: 267-268), Britain (1998: 215), Culpeper et al. (2018: 8) for a discussion of prosody and politeness.

information that may be quite intimate at times. The learners in the present study also share quite intimate information with the interviewer at times. Therefore, Britain (1998: 232) argues that uptalk is a positive politeness marker “used to emphasize speaker-hearer solidarity and to assist in the cooperative management of talk”. In sum, while learners may exhibit some L1 transfer features in the way they form their narratives, they may also adopt gender specific native-like prosodic patterns in how uptalk is used. A more fine-grained analysis of the different parts of the narratives instead of the simple dichotomies (Task 1 vs. Task 2, beginning vs. later parts of a recording) used in this study would be needed in future research.

In addition, uptalk in non-native speech also serves the purpose of preventing misunderstanding. Romero-Trillo (2007) and Romero-Trillo and Maguire (2011) developed the *Adaptive Management* model, which explains how conversations are managed through self-organization in order to avoid systematic breakdown. Romero-Trillo and Lenn (2011), who analyzed pragmatic markers and the lexical and intonational implications in cross-linguistic conversations, found there to be two types of pragmatic markers in Adaptive Management, overt and rhetorical. The overt ones mark something like “do you know what I mean?” which require a minimal response from a listener to verify comprehension. Rhetorical ones such as *you know*, *you see*, and *I mean* do not require a response from the listener, rather the correct cognitive reception is taken for granted and they serve to establish a rapport with the listener (Romero-Trillo 2002: 778, 2007: 84-85). I propose that uptalk in this study is also used to fulfill both of these functions, one to avoid pragmatic misunderstanding and the other to build rapport with the listener. In a corpus of L1 Spanish Romero-Trillo and Lenn (2011) showed that Spanish makes use of both types of pragmatic markers, overt and rhetorical (e.g. *explico* and *entiendes*), in order to prevent misunderstandings. By contrast, “[…] English does not favour the use of lexically transparent elements to indicate understanding” (Romero-Trillo and Lenn 2011: 232-233) but instead they seem to prefer the discourse markers *you know* and *you see* (addressee-oriented), and *I mean* (speaker-oriented) and in combination with the use of prosody to show differences in meaning nuances in these pragmatic markers, while Spanish speakers rather use formulaic variability (Romero-Trillo 2007: 89). Thus, Spanish and English differ in that their use of pragmatic markers and their intonation is rather a question of correct understanding than of accurate explanation (Romero-Trillo and Lenn 2011: 233). They also note that in English rising tones never realize the overt function of pragmatic markers (*I mean*), i.e. they do not invite/expect an answer, while Spanish speakers seem to prefer the overt function. In their qualitative analysis, Romero-Trillo and Lenn (2011: 238) found that NNSs of Spanish in conversations with NSs of English do not make use of Spanish strategies but rather try to mirror the usage of pragmatic markers of the NSs. I argue that the NNSs also use intonation to signal different pragmatic meanings, but they may not align them with pragmatic markers, but might produce them on other lexical items instead. The pragmatic markers themselves might be enough for them to signal the pragmatic meaning and especially when this is different in their L1, as is the case with the Spanish speakers, this might be an example of negative L1 transfer. Even in native speech uptalk has been compared to discourse markers such as *you know*, which can convey a “sense of speaker involvement” (Torres 2002: 69; Armstrong et al. 2016). Therefore, future studies need to include lexical features and pragmatic markers with overt and rhetorical functions in prosodic studies. Additionally, according to Romero-Trillo and Lenn (2011: 226), “Pragmatic misunderstandings occur in all types of conversations, also between native speakers of the same language”, and:

> […] it is interesting to notice how often speakers of the same language and background fall in the trap of pragmatic misunderstand even more easily than speakers of different languages and origins. In this case, we could argue that pragmatic misunderstanding is sometimes the result of habit or lack of attention rather than a question of cultural impairment or linguistic deficiency.

The native interviewers in LINDSEI were, however, very attentive and paid special attention to the meaning of the utterances of the interviewees. To conclude this argument, the learners in the present study may make use of uptalk for adaptive management in order to prevent misunderstandings, more overtly than NSs of English. The uptalk tunes be the SpE speakers were found to be among the steepest and highest. The use of uptalk by the Spanish learners in the present data set may be due to L1 influence, because the learners may feel like they have to make their meaning even more transparent when speaking in their L2 by overtly checking listener comprehension. All in all, the three L1s chosen in the present study have their own pragmatics of prosody. According to Elvira-García (2016: 42), the falling intonation for declaratives and rising for questions is a too simple distinction for Spanish. She contends that there are at least 14 different tones that convey different pragmatic functions (ibid.). Also in Czech, the confirmation tag *jasně* with different prosodic realizations (f0) lead to different pragmatic interpretations of the confirmation tag, from recognition to resignation (Volín et al. 2014, 2016). More research is needed on the prosody-pragmatics intersection, which will be valuable for further insights into L1 influence and L2 proficiency.

However, not every deviant feature can be seen as examples of negative transfer. The use of uptalk for confirmation checks may be attributable to the learners L2 learning experience, which mainly takes place in foreign language instruction classrooms (Mauranen 2006: 126), which may lead to the acquisition of certain fixed intonational patterns or "chunks" (Pickering 2009), i.e. fossilized intonational patterns with certain lexico-grammatical features to signal certain pragmatic functions, also called pragmatic fossilization (Romero-Trillo 2002). One of my hypothesis is that learners in this study acquired the rising end of the sentence intonation because they hear NSs interacting with them with rising and often exaggerated intonation (see also Knoll et al. 2006), i.e. they engage in foreigner-talk and they adopt that but use it to some degree in an inappropriate way. In other words, they produce high boundary tones more frequently than NSs in similar contexts (see also Romero-Trillo 2019). Interestingly, Romero-Trillo (2019: 96) found differences in the use of feedback (backchanneling) given by English NSs in the LLC and LINDSEI-SP. Not only do the NSs differ in their distribution of backchanneling devices but they also use different intonational patterns. The LLC speakers in the L1-L1 interactions use mainly a falling tone, while the NSs (the interviewers) in LINDSEI-SP (L1-L2 interactions) use a fall-rise and rise most frequently. Thus, the NSs seem to behave differently in the two corpora. According to Romero-Trillo (2019: 98) the NSs have different intonational patterns in their backchanneling because they are more emphatic when they speak to NNSs. In turn, the NSs in LINDSEI may also copy the prosodic patterns from the NNSs. Cogo and Dewey (2006: 70), for instance, state that lexical repetitions are often used in ELF discourse as an accommodation strategy in order to "signaling agreement and listenership and engagement in the conversation". The interviewers in LINDSEI are teachers and accustomed to giving feedback to learners that is stimulating and engaging in order to show listenership and engagement. According to Romero-Trillo (2019: 98), the learners engage in "prosodic matching" and copy the NSs' use of the pragmatic backchanneling devices.

> The fact that the range of markers and the prosodic patterning is different when L1 speakers interact with other L1 or with L2 speakers is a factor that can be considered in language instruction, and in tackling pragmatic fossilization, as we can conclude that L2 speakers are exposed to a different prosodic model that has fewer pragmatic nuances. (Romero-Trillo 2019: 101)

Another explanation may be that learners of English have been shown to transfer acquired prosodic regional variation to different speaking styles (Ulbrich 2008), which might also show up in their spontaneous productions, depending on their degree of proficiency and whether they are aware of different usages in more formal speaking styles. Thus, the learners may not be aware of using the uptalk feature inappropriately or using it too much, while NSs may be able to suppress this stigmatized variety in more formal contexts. Ulbrich (2008) contends that intonation used as a marker of stylistic variation may not be learned until later stages of the second language acquisition process. Therefore, it is not sufficient to have an L1 NS as a teacher, but it is also necessary to provide access to authentic L1 corpus data which according to Romero-Trillo (2019: 101) will be "essential to develop full pragmatic competence in the L2". Gass and Varonis (1985: 55): state that "comprehensibility of NNS speech is an important determiner of foreign talk", and foreigner talk "changes as a function of the NNS interlocutor's ability to understand and be understood in the L2" (56). In sum, the high frequency of rising tones in L2 speech may be explained by the fact that most of the prosodic models that learners are exposed to may be foreigner-talk and teacher-talk, which may exhibit a more unnatural and exaggerated rising intonation with less pragmatic nuances. My interpretation is that learners might use high edge tones more often to convey the same meaning because they have not been taught discourse markers as found for German learners by Götz (2013) and Spanish learners by Romero-Trillo (2002). More empirical studies are needed that investigate pragmatic functions of prosodic realizations in native and non-native speech in order to enhance language learning materials and teaching, so that learners do not only use one tag question with a fossilized tune but also learn different nuances that can be signaled by varying tunes on the same discourse markers and other parts of their utterances.

The results of the analysis of uptalk frequency and the stay abroad variable in the present study may lend support to the explanation that uptalk usage may be a fossilized feature acquired in the foreign language classroom. Since the analysis showed that learners with no stays abroad, short, and long stays abroad use uptalk equally frequent. However, the analysis has shown that a stay abroad in Ireland lead to an increased usage of uptalk. Additionally, the CEFR levels of the learners only showed extremely weak correlations with uptalk usage (positive in the case of the CzE, and negative correlation in the case of the GerE speakers). These results suggest that uptalk is a feature that does not seem to improve towards more native-like productions as a function of increased exposure to the TL. However, some learners approximate their TL, depending on the country of residence during a stay abroad, i.e. the Irish accent, where uptalk is a standard feature which is used in unmarked declarative utterances. To conclude, the use of uptalk in non-native speech can be primarily explained by developmental, proficiency-related factors. However, L1 influence may also play a role in the case of the Spanish learners.

5.5.4 Summary

The explanation of high boundary tones on declaratives has been discussed with regards to their pragmatic function, the uncertainty interpretation has been challenged and negotiation of meaning has been discussed. In the present study, uptalk has been identified as a polysemous pragmatic marker in non-native speech, with its different forms and different functions it can fulfill. In particular, uptalk in non-native English is also used to hold the floor (continuation), for checking (confirmation requests, checking comprehension, etc.), to signal uncertainty (in indirectly asking for feedback), and listing of items (similar to floor holding and at the same time checking for comprehension). Uptalk tunes in non-native speech can be interpreted to be used as at least four different pragmatic markers: as a marker of uncertainty, to signal new information, to indicate floor-holding, and an affective marker of in-group solidarity similar to pragmatic functions found in BrE and AmE (e.g. Levon 2016: 139; Burdin and Tyler 2018). The present study tried to show that uptalk is used for different pragmatic purposes similar to the ones found in previous research for better known and well-studied uptalk usage. Some of the uptalk patterns used by the NNSs in the present study thus fulfill a very similar function to the pragmatic marker *you know* in Buysse (2017: 54). In sum, uptalk in non-native speech has been found to fulfill three major functions which Pickering (2009) summarized: information functions, discourse management functions, and relationship-building functions (Pierrehumbert and Hirschberg 1990; Brazil 1997; Wichmann 2000; Wennerstrom 2001). Non-native speech, mainly female speakers from CzE and GerE, employed uptalk for these various pragmatic functions as interpersonal markers and to structure their discourse. These findings suggest that learners may go through a common developmental path in using uptalk in situations where NSs would not. Many explanations for the higher frequency of high boundary tones in general from L1 influence, to a lack of pragmatic accuracy and knowledge, and to interviewer influence (i.e. exaggerated foreigner-talk) and thus providing a different prosodic model to the learners from which they acquire their prosodic patterns have been proposed. In sum, uptalk in non-native speech was found to be used as a positive politeness marker, to reduce social distance, and it is very other-oriented, and is thus an indicator of proficiency (both higher and lower proficiency levels) rather than L1 influence.

5.6 SLA and prosody: Acquisition of L2 prosody

The findings summarized in Section 5.1 have a number of theoretical and empirical ramifications, which will be discussed in light with the SLA theories and predictor variables (previously discussed in Section 1.5) in Section 5.6.1. In addition, Mennen's (2015) LILt predictions are reviewed and a prosodic model based on the findings of the present study is presented (Section 5.6.2). The section closes with practical implications of the results for language teaching and speech technology (Section 5.6.3).

5.6.1 SLA-predictors (age, sex, semester abroad, etc.)

The present study set out to test if EFL learners' use of selected prosodic features can be predicted by their proficiency level/other factors (addressed in RQ3b). As explained in the theory, several extralinguistic factors can lead to favorable conditions for ultimate attainment, e.g. a younger AoA, a longer residence in a foreign-speaking country, and the length and type of instruction. The present section summarizes all predictor variables and their effect on the prosodic features (see Table 51 for a summary) and comes to a conclusion on the effect of these predictors. Concluding remarks on the contribution the present study made to SLA-theory and the acquisition of prosody in particular are given. Nevertheless, this section is very brief, because many SLA-related results were already discussed in the respective discussions by prosodic features and they will be thematized in the following two sections as well (5.6.2 and 5.6.3).

Table 51. Summary of all SLA-predictors by prosodic variable

Features	AGE	SEX	STAY ABROAD	PROFICIENCY	SPEAKING STYLE	FLUENCY
Tone freq. in phw	Older learners produce more tones	Females produce significantly more tones phw as compared to males	n.s. because no differences found between native and NNSs	CzE/GerE: -.09 negative correlation. The higher the CEFR level the fewer tones phw are produced	More tones are produced in both and dialogues as compared to monologues	The faster the IP in WPS and the longer the ip in seconds, the fewer tones are produced. The more UPs are produced the more tones are produced
High edge tones	n.s.	High tones are more frequent in female speech. The interaction SEX*INT_SEX shows that when both interlocutors are female, high tones are predicted to be used more frequently. Males retain their usage of high tones in both conditions, i.e. when they speak to a male or female interviewer	For CzE speakers an improvement can be observed (i.e. fewer high tones are used by learners with a stay abroad) but a contradictory development can be observed for the GerE speakers and virtually no change can be seen for the MadE speakers	CzE: -0.24 for high phrase accents phw and -0.33 for high boundary tones phw=negative correlation, i.e. the higher the CEFR level the lower the number of high phrase accents and high boundary tones GerE: 0.043 for high phrase accents phw and -0.16 for high boundary tones phw=weak correlations The more the learners have learned English in school and the more foreign languages they speak, the lower their production of high tones becomes	More dialogic speech (in post-interactions) more high tones are used	n.s.
IP freq. in phw	n.s.	Significantly more for GerE females, significantly less for SAmE and CzE females and MurE males	n.s.	-.27 (CzE) -.08 (GerE) negative correlation - the higher the CEFR level, the lower the number of IPs	More IPs are produced in dialogues and speaking styles containing both dialogues and monologues	More UPs and faster speech rate per IP lead to more IPs phw
IP speech rate in WPS per IP	n.s.	n.s.	n.s.	.63 (CzE) .31 (GerE) positive correlation, the higher the CEFR level, faster the speech rate. The higher the number of other foreign languages, the faster the IP speech rate becomes	Faster speech rate in dialogues	Slower speech rate the more EL, UP, FP, and HP are produced

Features	AGE	SEX	STAY ABROAD	PROFICIENCY	SPEAKING STYLE	FLUENCY
ip freq.	n.s.	CzE females and SAmE males are predicted to produce significantly fewer ips phw. SAmE female speakers are predicted to produce significantly more ips phw	n.s.	-.03 (CzE) -.10 (GerE) negative correlation - the higher the CEFR level, the lower the number of ips the longer the Czech learners were at university, the higher their ip frequency gets	n.s.	More UPs lead to more ips phw, in the learner data only, more FPs additionally lead to fewer ip breaks
ip speech rate in WPS per ip	<u>The older speakers are the slower their ip speech rate becomes</u>	n.s.	n.s.	.62 (CzE) .23 (GerE) positive correlation, the higher the CEFR level, faster the speech rate	n.s.	The higher number of fluency features such as EL, UP, FP, and HP all contribute to a slower speech rate
F0 level	n.s.	Men produce lower f0 level	n.s.	CzE: 0.043 - positive but extremely weak correlation GerE: 0.16 - same as for CzE	After post-interactions f0 level is raised	The longer the ip in seconds, the higher the f0 level
F0 span	n.s.	Men produce narrower f0 span	GerE speakers who have been on short stays abroad produce wider f0 spans for LH-tunes	CzE: 0.15 - positive but weak correlation. GerE: 0.093 - same as for CzE	F0 span wider in dialogues and after post-interactions	The longer the ips in seconds the wider the f0 span
Uptalk freq.	Older learners are less likely to produce uptalk instances	Female speakers produce significantly more uptalk instances	the learners who have been abroad to Ireland produce the most uptalking instances	CzE: 0.098 - positive but extremely weak correlation. GerE: -0.4 - negative correlation, i.e. the higher the CEFR level, the fewer uptalk instances phw are used	<u>More uptalk instances are produced in monologues</u>	n.s.

The results in Table 51 are based on the results including the native control groups as well as the results from the NNS-NNS comparisons. The variables STAB, CEFR, NO_OFL, YOEAU, and YOEAS were only investigated in the NNS models for the respective prosodic variables. Generally, all trends that were found in the model with the NS control groups were mirrored in the NNS models without the native controls. Only one age effect, four speaking style effects, and one interviewer sex effect were found in the models with the native controls but not in the NNS models, which were underlined in Table 51. The major difference between the two statistical models for each prosodic variable was observed for speaking style. The speaking style differences diminished when only the NNSs were compared to each other. This point will be discussed further below.

Age

The speakers' age did not play any significant role or only played a negligible role for most of the prosodic features investigated. This is perhaps due to the age differences between the speakers being too small and/or no age-based differences having been found. The only significant effects found were that older speakers produce significantly slower speech rates in ips, older learners produce more tones, and older learners are less likely to produce uptalk instances.

Sex

As expected, speakers of all varieties exhibited typical gender-based prosodic features, i.e. females produced significantly higher f0 levels, wider f0 spans, and differed mostly on the distributional level with producing more tones in general, and more high edge tones. Additionally, the non-native as well as native speakers demonstrated gender-based differences on the distributional level for IUs, with some female groups producing more or less IUs compared to males. The only prosodic feature that sex had no effect on was IU speech rate. Moreover, certain gender-pairings seemed to have an effect on high edge tone (more in female-female conversations) and uptalk frequency (more uptalk when interviewer is male).

Stay abroad

What all learners had in common was that the stay abroad had no effect on all but one prosodic variable, i.e. the frequency of high edge tones becoming less frequent for CzE speakers. The results of the stay abroad variable depending on the learners' gender and/or their L1 does not come as a surprise after consulting previous research (Section 1.5.2; e.g. Henriksen et al. 2010; Derwing and Munro 2013; Götz and Mukherjee 2018: 59). For uptalk instances, the length of the stay abroad did not matter but those learners who have been to Ireland produced the most uptalk instances. Therefore, demonstrating that the country of residence during a stay abroad also plays a role and needs to be considered in the analysis of prosodic features, especially uptalk. The conflicting results found (for GerE and MadE speakers), especially, may be explained by the fact that the learners have not spent sufficient time abroad. The CzE and GerE speakers have spent on average only 9 months abroad, while the MadE speakers spent only 5 months abroad. Previous research has shown that the L2 phonology may need more time to improve (e.g. Flege et al. 1997a; Flege 1988; Saito 2015; Flege 2016; Abrahamsson and Hyltenstam 2008, 2009; DeKeyser et al. 2010; Gut 2017).

Thus, not only longer stays abroad may be more beneficial, but also the intensity of interactions with NSs during a stay abroad may have to be strengthened and, ideally, L1 use may have to be reduced as well. Intonation is also partially resistant to natural exposure during a stay abroad but may still be acquired by learners who have not been abroad. A larger and more balanced data set with the stay abroad variable is needed for future research to obtain more conclusive results.

CEFR levels

The CEFR levels did not seem to be good indicators for most of the prosodic variables investigated, because mainly weak correlations were found. However, it has to be kept in mind that only few GerE speakers (n=19) and all CzE speakers were included, although they did not demonstrate a balanced number of all proficiency levels (B2-C2). The only moderate correlation found was for the improvement of IU speech rate for the CzE speakers with increasing proficiency. This suggests that while the CEFR levels may be good indicators for fluency, they do not seem to reflect on the learners' performance on other prosodic variables. Even though the correlations are weak, some tendencies for improvement can be observed, i.e. with increasing proficiency, the fewer tones are produced, the fewer IU prosodic breaks are made, and the fewer high edge tones are produced (CzE). Thus, just as Carlsen (2012: 162) summarizes, "levels of proficiency are not always carefully defined, and the claims about proficiency levels are seldom supported by empirical evidence", which is also reflected in fluency research by Dumont (2017) and Götz (2019b), who found difficulties in finding significant differences between the learner's performances and their rated proficiency level (i.e. CEFR level).

Speaking style

Speaking style was measured in two different ways, i.e. by task (mono vs. dia) and interviewer influence/involvement (pre- and post-interactions). Dialogues lead to more tones, more high tones, more frequent IP breaks, faster speech rate per IP, higher f0 levels, and wider f0 spans. The tonal features and f0 range results are mainly due to entrainment effects in post-interactions. There were no significant effects of speaking style on ip frequency and speech rate. IP frequency and IP speech rate were more speaking style-dependent than the same measures by ip. Additionally, uptalk was found to be more frequent in monologues. Pérez-Paredes and Sánchez-Tornel's (2019) study L1 interviews (LOCNEC) from a variationist perspective, making use of MD-analysis. They show implications for Language Testing and Assessment and learner corpus research, focusing on a critical perspective of task and test design. Their study showed that the LOCNEC/LINDSEI interview-format is a complex register, which can be classified as conversational language with a high dependency on task. All three tasks investigated differed from each other significantly on almost all dimensions. One major argument the study makes is that the different speaking tasks in LOCNEC yield different sub-registers. In the present study the speaking style did not seem to be a big influencing factor as first assumed. As in previous research, the differences between monologic and dialogic speech are negligible (cf. Jiránkova et al. 2019: 23), since speaking style always returned higher p-values for all models except for IP frequency, where the speaking style effect was highly significant. In terms of speech rate per IP there is a significant increase in WPS from monologues to dialogues and a non-significant increase for ips, indicating that dialogic speech is slightly faster than monologic speech. This can be explained by the higher informational load and structural and lexical complexity in monologues that require more planning time (cf. Jiránkova et al. 2019: 23). In sum, while speaking style differences were observed on all prosodic variables (except the ip level), the two speaking styles analyzed exhibited similar prosodic productions and entrainment effects were found in post-interactions. When NNSs are compared to each other, only minor speaking style differences are observed due to fuzzier boundaries between dialogic and monologic speech in non-native speech.

Fluency

The more dysfluencies are produced (EL, UP, FP, and HP), the slower the speech rate per IU becomes. The more silent pauses are used, the more IUs are produced. And the faster the speech rate, the more IPs are produced. The faster the IP in WPS, the more tones are produced. The more silent pauses are produced the more tones are produced. The longer the ips in seconds, the fewer tones are produced. In the learner data only, the more filled pauses are produced, the fewer ip breaks are made. Moreover, longer ips in seconds lead to higher and wider f0 range. In sum, pausing behavior and other dysfluencies explain most of the distribution of IU frequency, IU speech rate, tone frequency, and f0 range measures, but did not explain the variation for high edge tone and uptalk frequency. This suggests that combining fluency variables with prosodic ones is beneficial for most prosodic variables.

Proficiency

Proficiency was tested in many different ways in the present study. The variables were lexical complexity (TTR, Guiraud-Index), CEFR levels, age, length of stay abroad, fluency variables (filled and silent pauses, WPS per IU, elongations, and hesitations), and years of English at school and university. None of the lexical complexity measures proved have a significant effect on any of the prosodic variables investigated. For most prosodic variables, no improvement after a stay abroad or higher CEFR level can be found, although fluency is the most reliable indicator to the production of most prosodic variables. While fluency features in the present study explain most of the variation, the stay abroad variable only explains the production of high edge tones, and the CEFR levels only show an improvement of IU speech rate for CzE speakers. While Götz (2019b) showed that the use of filled pauses decreases with increasing proficiency (Spanish learners among others), however, the low explanatory power of the regression analysis points towards previous findings that proficiency alone is not a good predictor of pausing behavior, and that country of origin and age are better predictors. Furthermore, performance phenomena have been shown to correlate among other factors with the relationship to the hearer (i.e. level of familiarity, e.g. Götz 2019a). Years of English at school only showed significant effects for the production of high tones, which becomes lower with an increasing number of years. Years of English at university only had an effect on the number of ip frequency, which becomes higher with an increasing number of years. Taken together, fluency explains most of the variation while the stay abroad variable and CEFR levels do not seem to be good predictors of prosodic behavior. However, combining various fluency variables is a rewarding endeavor, since proficiency is a complex construct that is influenced by various factors.

Number of other foreign languages

The sheer number of other foreign languages spoken by the NNSs in the present data set did only have an effect on two prosodic features analyzed, i.e. the more foreign languages the learners speak, the lower their production of high tones and the faster the IP speech rate is predicted to be. However, the specific foreign languages (French, German, etc.) were not included in the analysis and information on the proficiency of these foreign languages is unknown for the present data set. Future studies may include this variable for more conclusive results.

Summary

To conclude, it can now be said that all predictors analyzed contribute together to the shape of L2 prosody of the three learner groups investigated. On the theoretical conceptualizations of L2 prosody, the results suggests that including only one measure of proficiency may not give a complete picture for all prosodic features. Each predictor sheds light on a different prosodic feature and intonational dimension. This will be discussed further in Section 5.6.2, where a multifactorial model will be proposed that tries to explain the results of the present study. This has implications for language testing and assessment and CEFR descriptor scales in particular, and therefore this will be discussed in Section 5.6.3, among other language-pedagogical implications. Overall, some of the assumptions of the LILt model (at least those that were testable with the present data set), were substantiated by the corpus analysis. A more detailed discussion will be provided in the following section.

5.6.2 Intonational dimensions and LILt predictions

The present section discusses the factors that determine the acquisition of L2 prosody and how learner language in general develops. In the development of learner language, there are different factors that affect learner language, i.e. the specific L1s the learners speak, their ELT experience and NS models, exposure to varieties of English, and their personal preferences, i.e. the variety of English they aspire to, among other factors. In this regard, I aim to place my findings and discuss them within Mennen's (2015) LILt model. Mennen's (2015) LILt incorporates four intonational dimensions along which intonation can differ from language to language and includes several assumptions similar to the SLM and PAM-L2 model (see Section 1.4.3). Specifically, the following research questions posed by Mennen (2015: 184) will be taken up in the present section:

1. *Are some intonation parameters more susceptible to transfer than others?*
2. *Are deviations equally reflected in different dimensions of intonation?*
3. *Do deviances in different dimensions of intonation diminish in parallel?*
4. *Are there symmetries in the pace and trajectory across learners of different L1 backgrounds?*

How the learners in the present data set behave in relation to the four LILt dimensions and in comparison to the NSs in the four prosodic features investigated are summarized in the prosodic model in Figure 50. Figure 50 depicts factors (see SLA predictors in Section 5.6.1) that generally affect learner language and relates this to the situation of the three learner varieties in the present study.

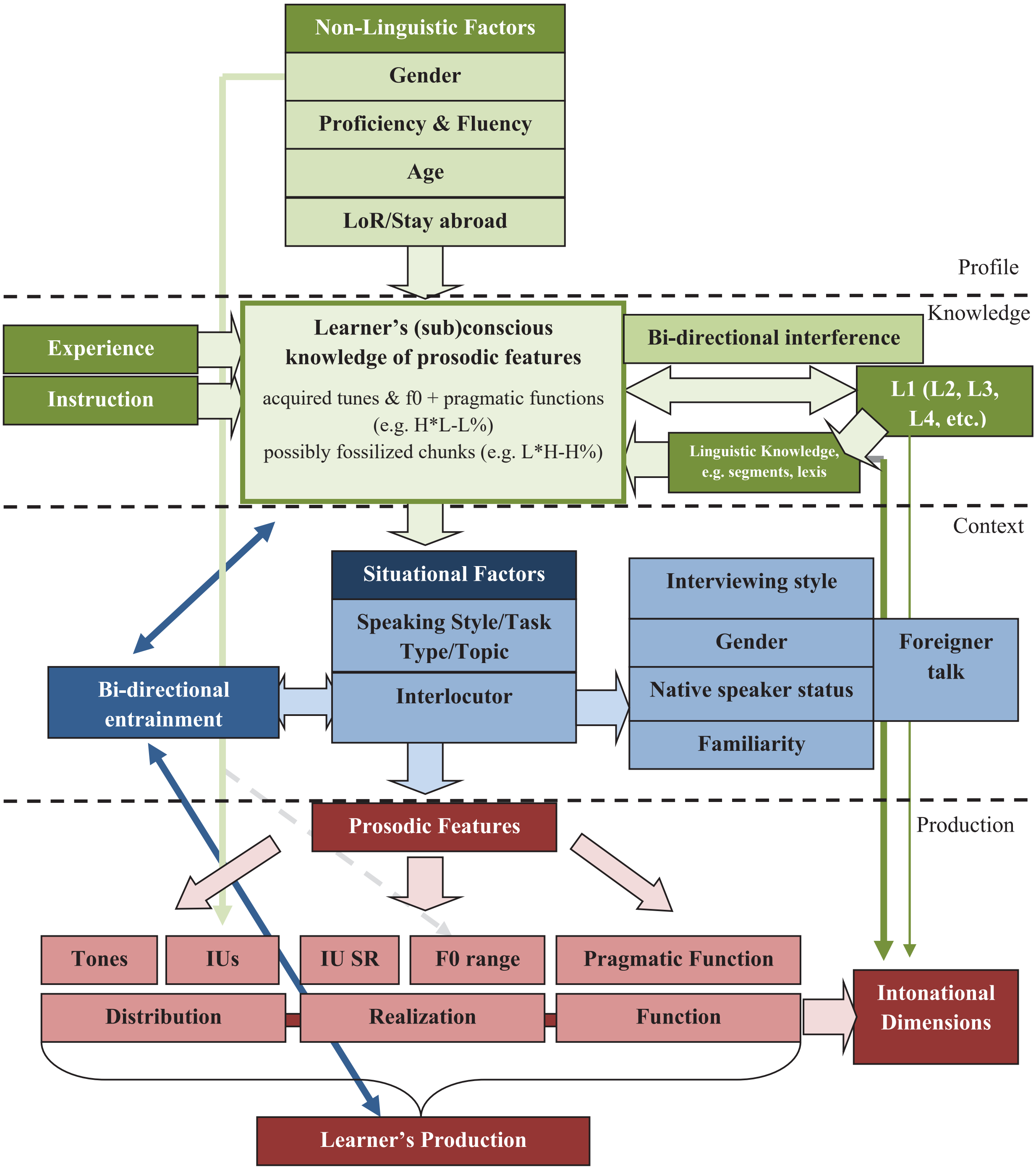


Figure 50. Explanatory model of the prosodic productions of the Czech, German, and Spanish learner groups (SR = speech rate)

Which prosodic features a learner produces depends on four dimensions (see Figure 50). The model suggests four interwoven layers that characterize the structural and functional make-up of the prosodic productions of learner language. First, the profile, i.e. the predisposition of the learner, needs to be taken into consideration. The learner profiles were used in the present study as the baseline for comparison across learner varieties and in reference to a norm-providing variety, i.e. NBrE (and AmE). The prosodic productions are constrained by a speaker's gender and age and his/her overall proficiency and experience. The learner's gender has the biggest influence on the distributional level of the intonational dimensions, since large gender-specific differences were found on the distribution of intonation unit frequencies. Naturally, gender also had a great influence on the realizational level (f0 range), which was expected due to biological as well as social reasons, while the IU frequency differences rather seem to point towards social and discoursal differences between the two sexes. The learner can only produce prosodic features that are within his/her physiological limits (sex, age) and learned behavior (socially accepted norms of f0 range for a specific gender). The non-linguistic factors in the model in Figure 50 are sorted according to their importance based on the results in the present study. Gender and fluency-related features (which I view as part of the learners' overall proficiency) were determined to be among the most important factors to explain L2 prosodic variation. The age and stay abroad play a rather marginal role in explaining prosodic variation in the learner data.

The second section concerns the learner's knowledge, which can involve (sub)consciously acquired tunes and f0 with or without their pragmatic functions, as well as possibly fossilized chunks along with their prosody. This knowledge is further influenced by the learner's experience with NSs and the language in general, (prosodic) language instruction, as well as the first and other languages acquired by the learner. The later can also be influenced bi-directionally, i.e. the learner's prosodic knowledge of English may transfer to his/her L1. Other foreign languages (e.g. L3, L4, etc.) a learner speaks/knows are included in the model because the results revealed that an increasing number of other foreign languages lead to more native-like high tone frequencies and IP speech rate. The linguistic competence (e.g. segments and lexical repertoire) of a learner was included under his/her knowledge, which also have an influence on which prosodic features are acquired/produced, and which in turn are possibly influenced by the learners' L1 and overall proficiency in the L2. For instance, elongations of segments had an influence on the production of IU boundaries.

The third section includes the context in which the learner finds him-/herself. These situational factors analyzed in the present study are speaking style/ task type and the topic the learners talk about. The interlocutor exerts the biggest influence on the learner's productions, because bi-directional entrainment can take place, i.e. the learner imitates the interlocutor and vice versa. To what extent the interlocutor influences the learner's prosodic productions further depends on the interlocutor's interviewing style, gender, NS status (and whether he/she engages in foreigner talk), and the level of familiarity between the interlocutors. Entrainment effects were mostly visible on the f0 range level and speaking style differences on the IU level. The model tries to demonstrate that f0 range entrainment effects may not only have an immediate and short-term effect on the IUs as a result of an interaction between interlocutors. The prosodic patterns observed may very well also be or lead to fossilized features in the learner's knowledge, which may have been stabilized through regular and repeated contact to foreigner-/teacher-talk by native speakers they interact with in a classroom setting or during a stay abroad. Therefore, the context, i.e. whether a native or non-native interlocutor is present, plays a large role on all levels into all directions, i.e. bi-directional entrainment on the interviewer's and/or interviewee's prosody is possible, the learner's knowledge may possibly include memorized and fossilized chunks, as well as immediate effects on the learner's final output during an interaction can also be expected.

Finally, the fourth section shows that the different prosodic features a learner (L1 Czech, German, or Spanish) produces may differ on mainly three out of four intonational dimensions, i.e. distributional, realizational, and the pragmatic function (cf. Ladd 2008; Mennen 2015). This corroborates previous research that showed that the realizational and functional dimensions of intonation seem to cause difficulties for most learners, while the systemic dimension does not seem to be an issue (cf. Mennen 2015; Ward and Gallardo 2017: 22). The arrows in between the three dimensions in Figure 50 indicate that the intonational dimensions may influence each other, which will be discussed further in the following paragraphs. In sum, a learner's final prosodic production is ultimately influenced by all three intonational dimensions differing by L1, linguistic competence on other linguistic levels, as well as gender. The following paragraphs specifically address Mennen's (2015) research questions.

Mennen's (2015: 184) first research question asks whether some intonational features may be more susceptible to transfer than others. As in the present study, the realizational dimension is often found to differ the most from the target norm, while the systemic dimension does not seem to cause many problems, which is often explained by an L1 influence (Gut 2009: 248; Weingartová et al. 2014: 236; Mennen 2015). While it is entirely possible for learners to produce prosodic output that is within the L2 norms, merging effects (cf. De Leeuw 2012; Mennen et al. 2014; Volín et al. 2015; Zárate-Sández 2015, 2018) as well as "overshooting" have been found (cf. De Leeuw et al. 2012). Since the present data set did not include L1 recordings of the respective learner groups, a direct L1 influence cannot be inferred. However, some of the results corroborated previous research on prosodic transfer phenomena from CA. Based on the literature review, it was hypothesized that the three learner groups would exhibit negative transfer in f0 range and positive transfer of the tone inventory. Due to the rather syllable-timed nature of Czech and Spanish, negative transfer was also expected in intonational phrasing. A possible L1 influence was proposed for all prosodic features investigated, with L1 group differences being found as well. Generally, however, since several similarities were found between the three learner groups from different L1 backgrounds, proficiency-related factors were deemed more likely to explain the results than L1 influence. In the following, the possible transfer phenomena discussed in previous sections (5.2-5.5) are summarized:

- **Tones and tunes:**
    - **Overproduction of L* H- H% tunes in CzE and GerE**: Czech and German speakers tend to start their utterances rather low and expand their f0 range in later parts of utterances (Mennen et al. 2012; Chamonikolasová 2017), which was confirmed for the CzE and GerE (also for SpE speakers) productions in this study.
    - **Wider variety of high pitch accents in CzE:** Slavic languages have been found to produce a more variable f0 compared to Germanic languages

(Andreeva et al. 2014a, b). The CzE speakers in the present study were found to use a wide variety of phonetic variants of simple high tones (e.g. !H*, ^H*, H*).

- **Overproduction of high pitch accents in SpE:** While previous research determined that SpE speakers also produce a wider variety of pitch accents compared to English NSs in previous research (cf. Ramírez Verdugo 2006a: 22) and an L1 interference of late peaks is expected (cf. Méndez Seijas 2018), SpE speakers in the present study were only found to overproduce high pitch accents compared to NBrE speakers, which may also be explained by an L1 influence (see also Ramírez Verdugo 2002: 122).

- **IU frequency:**
  - **Underproduction of IU breaks in CzE:** CzE speakers produced the fewest IU breaks, the slowest speech rate in ips, the most elongations and filled pauses, more hesitations, but produced a native-like frequency of silent pauses. In the present study, it was determined that silent pauses and a faster speech rate functions as disjuncture, while filled pauses function as IU-binding features. Therefore, due to L1 transfer of fluency phenomena (elongations, filled pauses, IU speech rate, and hesitations), the CzE speakers produce fewer IU breaks.
  - **Native-like IU frequency in SpE:** Although SpE speakers have also been shown to transfer fillers and elongations from their L1 into the L2 (Clark and Fox Tree 2002: 93; Ortega-Llebaria and Colantoni 2014; García Lecumberri 2017: 184), in the present study they did not exhibit any deviances from the NS norm on the IU frequency level despite a frequent use of filled and silent pauses, as well as hesitation phenomena (applying more to MurE speakers than MadE, who produce more native-like fluency phenomena). However, if fluency phenomena are not taken into account on the ip level, SpE speakers are predicted to produce significantly more ips than NBrE speakers.
- **IU speech rate:**
  - **Slowest speech rate in CzE and MadE:** CzE speakers were found to transfer their speech rate, filled pauses, phrase-final lengthening to mark prosodic boundaries, and repeats from their L1 into the L2 (Dankovičová 1998; Volín and Skarnitzl 2007; Jiřelová 2018; Jiránkova et al. 2019). In the present study, CzE and MadE speakers produced the slowest speech rates within IUs, which may be due to the syllable-timed nature of the respective L1s, as well as the increased use of filled pauses, elongations, and hesitations (including repeats) in the case of the CzE speakers.
- **F0 range:**
  - **Narrow f0 range for HL-tunes in all learners:** Czech, German, and Spanish have been described to have a lower f0 range compared to English (e.g. Kelm 1995; Scharff-Rethfeldt et al. 2008; Estebas-Vilaplana 2014: 185; Volín et al. 2015: 111). This was found to be true for all non-native varieties in the present study but only for HL-tunes.
  - **Wider f0 range for high-register tunes in all learners:** However, f0 range was found to be expanded in later parts of an utterance, which was partly

explained by an L1 influence (cf. (Mennen et al. 2012; Chamonikolasová 2017).

- **Narrowest and widest f0 range in MurE:** MurE has been described to be among the flattest varieties of Spanish accents (Monroy and Hernández-Campoy 2015), which might explain why they deviate the most on the f0 range level, i.e. narrowest f0 span for HL-tunes (L1 influence) and widest and highest f0 for all other tunes (prosodic drift/avoidance of L1 patterns).

One of the LILt assumptions was that increasing experience in an L2 will lead to more native-like production of L2 intonation, and learners with limited experience will rely on their L1 structures so that L1 transfer will take place. This claim could not be substantiated in the present study, since all learners showed possible L1 transfer phenomena, regardless of their proficiency level. From the summary above, it seems that f0 range seems to be more susceptible to transfer than the other prosodic features, since an L1 influence can be observed for all three L1 backgrounds (i.e. a narrower f0 range for HL-tunes). IU speech rate may be more susceptible to transfer for learners with more syllable-timed L1s (i.e. Czech and Spanish), which also may cause transfer effects on the IU frequency level for CzE speakers but not for SpE ones. Additionally, Slavic languages have been shown to produce more variability in f0, which possibly lead to less extreme f0 range for high-register tunes in CzE speakers, who used a variety of phonetic variants of high tones (H, !H, and ^H). SpE speakers, on the other hand, produced fewer high-register tunes in general but they did with less phonetic variation, i.e. they consistently reproduce extreme f0 values (^H). In sum, my hypotheses posed in the theory have been confirmed, i.e. there seems to be negative transfer of f0 range, positive transfer of the tone inventory, and negative transfer of intonational phrasing behavior (particularly IU speech rate) can be assumed for more syllable-timed languages. There also seems to be a tendency for CzE speakers to be the most susceptible to L1 transfer, which is visible the most on the IU level. This may be due to Czech being the most typologically different language from English, as compared to German and Spanish. Nevertheless, it has to be kept in mind that all learners in the present study deviate on all prosodic levels in the same direction (except for IU frequency), although with different degrees of strength of the deviances and, therefore, L1-based explanations may be secondary.

To answer research questions two and three by Mennen (2015: 184), whether deviations are equally reflected in different dimensions of intonation and whether they diminish in parallel, it can now be said that deviances on three out of four intonational dimensions were found. However, the number of prosodic features analyzed differed by intonational dimension, i.e. five distributional (IP, ip, tone, tune, and uptalk frequency), four realizational (f0 level, f0 span, IP speech rate, ip speech rate),[81] and one functional dimension (pragmatic functions) were investigated. Despite the imbalance of prosodic features analyzed per intonational dimension, deviances for most prosodic features on all three dimensions were found. Overall, however, the realizational dimension seems to cause the most difficulties for all learner groups. Furthermore, L1- and gender-based differences were found on which intonational dimension learners tend to differ more. On the distributional level, the most

[81] Plus five uptalk features, which were only analyzed for CzE and GerE speakers: uptalk f0 span, uptalk f0 level, uptalk f0 min, uptalk rise onset, and uptalk slope of rise.

commonly used tune is H* L-(L%) and the tune L* H-(H%) is used more by non-native and female speakers. On the IU frequency level, gender-based differences were also observed (distributional dimension), with CzE and GerE females using different intonational phrasing strategies. Therefore, the light green arrow in the model in Figure 50 demonstrates that the distributional level of intonation is mostly influenced by the learner's gender. Besides gender-based differences, also L1-based differences were found, with SpE speakers exhibiting more difficulties on the realizational level (f0 range), and CzE and GerE (mostly female) speakers deviating more from NSs on the distributional level (IU frequency and high edge tone frequency), and consequently also on the pragmatic level, due to higher frequency of high-register tunes. The dark green arrow in Figure 50, therefore, indicates that the intonational dimensions differ depending on a speaker's L1. Differences in distributional and realizational dimensions indicate that the learners have not reached ultimate attainment for all prosodic variables yet. The pragmatic functions and tonal choices are rarely analyzed in non-native speech and seem to be a problem even for advanced learners of English. Toivanen (2003) found pragmatic rather than purely phonetic interference to be more significant, since no substantial differences in the distribution of tones were found. No matter what the pragmatic function of the utterance, the learners always used a falling pattern, while the British NSs used a fall-rising pattern to signal different pragmatic functions of IPs. In the present study the NBrE speakers were found to use a HL-tune to mostly signal continuation in monologues and handing over of turns in dialogues, while the learners frequently also use the reverse LH-tune (and HL-tunes) to signal a variety of pragmatic functions. While the learners do not misuse prosody on the pragmatic level, they were found to produce more different tunes and f0 range to achieve a pragmatic effect, which were mainly explained by a need to make their pragmatic intent overly explicit. The NSs may not need to make their pragmatic intent overly explicit, which also explains the distribution of simple H* L-(L%)-tunes to signal continuation and handing over of turns. Especially the pragmatic functions of checking and signaling of uncertainty were rarely used by the NSs. Further inferences on the functional level cannot be made because only one variable was analyzed, which was conditioned by different tonal choices and f0 range. Thus, the learners only deviated on the functional level due to deviances on the other two levels. This shows that all prosodic features and the different intonational dimensions are all dependent on each other, i.e. a deviance on one level may lead to further deviances on another. For instance, the GerE female speakers produce more IPs (possibly due to their fluency/proficiency level), and, therefore, have more of a pragmatic need to signal cohesion, continuation, etc. and produce more high-register tunes with a wider and higher f0 to make their pragmatic intent overly explicit. Thus, the deviances on the IU frequency level lead to deviances on tonal frequency and choice, as well as f0 range and pragmatic intent of these tonal patterns. It does not have to necessarily start with IU production difficulties but instead the pragmatic intent could come first due to specific IU and tonal choices that make the L2 output more comprehensible. In sum, while deviances on all intonational dimensions (three out of four) were found, this is conditioned by an interdependency of the dimensions, and the extent of deviances on the three dimensions further depends on the L1 and gender. The variational investigation of prosodic phenomena also confirms one of the assumptions of the LILt model that not all intonational dimensions constitute the same amount of difficulty for all learners and that ultimate attainment is possible in some intonational dimensions (for some prosodic features), more so for the distributional than the realizational one for advanced learners of English. The results provide some indication for variety-specific preferences for intonational phrasing frequency, which is the prosodic feature on which the learners differ the most from each other.

RQ4 (*Are there symmetries in the pace and trajectory across learners of different L1 backgrounds?* (Mennen 2015: 184)) was partially answered in the previous paragraph in which L1-based differences and similarities were discussed. These similarities in deviations in the interlanguages with different L1 backgrounds represented in the prosodic model above appear to indicate that the learners share universal developmental paths in acquiring the prosody of an L2. This is in line with previous research that also observed universal developmental paths for the acquisition of L2 prosody (e.g. Archibald 1994). More specifically relevant to SLA-theory is that the analysis suggests that prosodic features may be learned in a specific order. Based on what all learners have produced in a native-like and non-native-like manner in the selected speech samples, as well as L1-based differences, and in consideration with the analyzed SLA predictors, an attempt has been made to sort the analyzed prosodic features according to their order in which they may be acquired during the SLA process. The results suggest the following order for CzE and GerE speakers: Tone frequency, f0 level of HL-tunes, IU speech rate and length (more deviant for CzE), tonal categories (more deviant for GerE), f0 range (mostly span), and IU frequency may be acquired last (by female speakers of both groups). For SpE speakers the following order is suggested: Tone frequency, f0 level of HL-tunes, IU frequency, IU speech rate and length (more deviant for MadE), tonal categories, f0 range (mostly span, more deviant for MurE speakers) may be acquired last. Tone frequency and f0 level of HL-tunes seem to be acquired without any problems because these are the only prosodic features all learners produced in a native-like manner. Therefore, one could assume that these features are acquired first and with relative ease for all learners with different L1 backgrounds. The SpE speakers additionally produced native-like IU frequencies and, therefore, the order of acquisition may be different. Besides this difference, the learners all follow a similar developmental path and show systematic difficulties with choosing tonal categories (edge tones), IU speech rate and length, and f0 range. The choice of tonal categories goes hand in hand with f0 span variation, since the learners have been shown to start their HL-tunes lower than NSs. They, therefore, produce an overall narrower f0 span. The production of HH- and LH- tunes leads to the opposite results, i.e. by choosing this tonal category learners produce wider f0 spans at the end of utterances as compared to NSs. Due to this logical interdependency of these features, tonal categories and f0 range may be acquired simultaneously. Finally, there are L1-based (and also gender-based) differences between the learner groups, which may suggest that these are the most difficult prosodic features to acquire for some speaker groups: IU frequency and high edge tones, which mostly affect the more advanced CzE and GerE female speakers and f0 range and tonal category choices for MurE speakers. Taken together, this may indicate the order of acquisition of prosodic features, i.e. tone frequency and f0 level of HL-tunes may be acquired first, while the tonal choice and accompanying f0 span, as well as IU speech rate and –frequency may be acquired at higher levels and would need more specific attention and targeted instruction (applies to CzE and GerE speakers). Since the MurE speakers who have been described to be the least proficient of all learners (my perception), have been shown to have difficulties the most with f0 span, this may be taken as a prosodic feature that is learned later on in the SLA process for this L1 group. The CzE and GerE speakers have been described as more advanced than the MurE speakers (based on available CEFR level ratings), and the fact that these two groups have difficulties with the same prosodic features suggests that these may be features to focus on with more advanced learners, i.e. IU frequency and tonal categories (edge tones). For the CzE speakers high edge tones became less frequent with a stay abroad and, therefore, for this group IU frequency should be a priority in language instruction, because no improvement for this variable was found. Therefore, as will be discussed in Section 5.6.3, it may be most beneficial to focus on intonational phrasing and tones for more advanced learners (CzE/GerE) first, and on tones and f0 range for lower proficiency learners (SpE). Thus, the present study lends further support to the hypothesis of modularity in SLA, which contends that the various linguistic levels (e.g. fluency, lexis, prosody) are learned independently from each other (e.g. Gut 2009; Skarnitzl and Rumlová 2019), which also applied to a more fine-grained level in the prosodic domain. Another widely held assumption in SLA theory is that ultimate attainment in phonology is not possible after a certain maturational level (i.e. age cut-off point) and that learner language ends in fossilization. The present study adds to these SLA-theories that ultimate attainment is possible for certain prosodic variables (IU speech rate and high edge tones in CzE; IU frequency in SpE, tone frequency and f0 level for HL-tunes for all learners, and f0 entrainment for CzE and GerE speakers), yet only from a productive and synchronic perspective. However, fossilization is difficult to prove and more research is needed to substantiate any claims made in the present study. Since none of the prosodic variables showed an improvement after a stay abroad (except high edge tones for CzE speakers), it is likely that all prosodic features are fossilized. Furthermore, for the CzE speakers a moderate and positive correlation was found for IU speech rate with higher CEFR levels, which may suggest that fluency-related features improve, while all other prosodic variables do not. In sum, the results suggest that learner language is quite systematic in terms of prosodic production across L1 backgrounds and, therefore, shows universal features, with the three intonational dimensions differing equally from each other, but with different L1 speakers deviating more (i.e. more extreme values but still in the same direction) than others on certain dimensions.

To conclude, I offer a few remarks on the BV-theory and MDH. The prosodic patterns used by the learners in the data set of the present study are far from being rudimentary or being "basic" (BV). For instance, the intonation used by the learners fulfills a variety of pragmatic functions, it marks a transition to reported/quoted speech, and the learners entrain to native speech. From the viewpoint of interlanguage intonation, these findings show an advanced level of prosodic features that seem to be universal to the three learner groups with different L1 backgrounds. The finding that uptalk features were used similar to native usage (reported in other studies e.g. Levon 2016, 2018) can be taken as a confirmation of one of the assumptions of the LILt, which proposes that the more dissimilar an L2 prosodic feature is, the easier it will be to learn. Maybe that is why learners use uptalk because it does not exist in their L1 (see also Flaig and Zerbian 2016) and they may have picked it up in their semester abroad or from foreigner speak used by their language teachers. Future research will have to determine how universal markedness (Rasier and Hiligsmann 2007) affects the ease of acquisition of L2 prosody. A case in point in the present study is that of the learners who used uptalk, which is a marked feature in NE, which is not used in the learners' L1s. According to the MDH it should be more difficult to acquire because of the features markedness status. However, as the analysis showed, some learners do use the feature in a native-like manner with pragmatic functions that are clearly distinguishable from the signaling of uncertainty. The implications of these results will be discussed in the following section.

5.6.3 Select implications for ELT

The findings of my research have considerable implications for language teaching and learning. The present section discusses some language-pedagogical implications derived from the findings and provides an outlook for future developments in prosody training and speech technology. However, some initial considerations about prosody research in general have to be made.

> Practicing pronunciation without prosody is like teaching ballroom dancing, only the students must stand still, practice without a partner, and without music. Gilbert (2008: 9) citing a teacher in training after a training course.

> So it appears that pronunciation practitioners never got the message that pronunciation was Cinderella no more, and the metaphor has continued to have power over how we see ourselves: victimized, neglected, and left at home when everyone else goes to the language teaching ball. (Levis 2019: 1)

> In the words of Tracey Derwing, we are no longer Cinderella but have instead become “the Belle of the Ball”. (Levis 2019: 6)

Levis (2019) argues that the victim mentality of pronunciation practitioners needs to stop, since the discipline has grown over the past decades, publication numbers have increased, and there are specialized journals and conferences as well as several software and web pages. However, Levis claims that we need more research and more material and especially more teacher education. While this is true for pronunciation, there is also a similar trend in the teaching of prosody. There is an increasing interest in the research community on L2 prosody and there is a growing body of research and material on how to teach prosody, making the ballroom metaphors less applicable to the field today. The following now turns to a discussion concerning the implications the findings of the present study provide from an applied linguistic perspective. The focus is on language teaching, i.e. teacher training, the development of teaching materials, teaching tips on which prosodic features to include and focus on, and how to teach them, along with language assessment. The present section is by no means an exhaustive discussion of the implications for ELT but focuses on a few select implications derived from the findings. The rest of this section is structured as follows, first opportunities and factors that determine successful L2 prosodic attainment are discussed, then the importance of determining and teaching the reference accents learners try to aspire to is discussed, teaching (high) tones in general, teaching materials that are available, including prosodic features in the ELT classroom, some notes on language assessments and the CEFR descriptors are provided, and teacher training is addressed in light of the findings on L2 prosody. The section closes with a few notes on implications in reference to speech technology.

While the first aspect to be discussed is mostly beyond the control of teachers and learners, one major aspect of successful L2 prosodic attainment is the geographical location where the learners reside, the quality of ELT in larger cities and smaller regions, and the access they have to NSs either through tourism, foreign residents and speakers of other languages in their own countries or through travel opportunities to English speaking countries (Jakšič and Šturm 2017). What the Czech Republic[82] and Spain have in common is that ELT does not have as strong a tradition as, for instance, in Germany, although efforts have been made in both countries in TEFL to raise average proficiency levels in English (Education First 2011-2015: 6; Carrie 2017). One aspect observed in the productions of the SpE learners is that the Madrid speakers have more native-like prosody than their Murcia counterparts. One could argue that the capital city may have more favorable conditions for ELT and language learning opportunities outside the classroom, which could partially explain the difference in performance. There is evidence that students from capital cities travel more, and they may be encouraged more by their parents and teachers to do so than students from regional schools (Jakšič and Šturm 2017; Czech Statistical Office 2019). In addition, the purpose for a stay abroad may affect the prosodic production of the learners. While most of the GerE and CzE speakers spent a semester abroad at a university, the MurE speakers mostly talk about a short aupair stay or other work and travel stays in English speaking countries. Therefore, their context of acquiring English may have included different experiences, and this difference needs to be accounted for in future research. Overall, the present study and previous research have shown that a stay abroad alone does not always contribute to native-like prosody. Therefore, study abroad programs should pay more attention to the type and amount of NS input learners receive by, for instance, promoting more study groups or other activities that involve other NSs as well and not only NNSs where ELF is spoken. This should be coupled with specific prosody instruction during the time abroad, before and after as well. Collaborative, information-exchange activities to create more opportunities for learners to engage in conversations with their classmates and to negotiate meaning (cf. Pica 1987) are also needed.

Which variety of English the learners try to approximate to cannot be determined with the present data set. However, which variety of English the NNSs come closer to in their productions can be, however, determined to a certain extent. All statistical models in the present study took the NBrE corpus as the intercept, for multiple reasons. The initial reason was due to the format of the dialogues, which resembles the format from the learner corpora the most. Previous research has also found a clear preference of learners in Europe for the BrE accent and easier recognition of the accent (Jakšič and Šturm 2017: 353; Skarnitzl and Rumlová 2019). Citing previous research, Culpeper (2018: 18) contends that "some people use English not to identify with native English speakers, but to achieve mutual understanding with other speakers". Even though the encounters with BrE in real life, i.e. through stays abroad or through tourism by BrE speakers is more likely (see Mompeán González 2004; Jakšič and Šturm 2017: 360) in all three countries (Cz, Ge, and Sp), the increase in the exposure to AmE culture in these environments should not be underestimated (cf. Jakšič and Šturm 2017: 356). The trend might shift more towards AmE as a reference accent. The learners in the present study showed differences based on their exposure and thus their

[82] English was only recently established as a main foreign language in the Czech Republic. Before 1989, Russian was the main foreign language. After that, it was German and then English.

possible preferences for a reference accent. The reference accent should be recognized and addressed in language teaching:

> Learners of English need to be prepared for extensive variation in the intonation they might hear from native speakers, within and across dialects […] Most of all, they need to be aware that variation in the southern 'standard' is as high or higher than in northern varieties of English spoken in the British Isles. In other words, the standard variety is no more uniform than non-standard varieties. (Grabe et al. 2005: 331)

Future teachers, therefore, have to be trained to distinguish between prosodic differences between different varieties of English before they can teach prosody, because some learners may not want to sound British or American and would therefore not benefit from or approve of such lessons. When teaching prosody as well as developing materials for language teaching, some aspects of prosodic change and the target model have to be kept in mind. More recent investigations of modern BrE show that there are linguistic changes on the segmental and intonational level. In his investigation of modern BrE, Lindsey (2019) contends that BrE is becoming more similar to AmE pronunciation. However, only certain intonational features seem to be prone to change.

> Neither phrasing nor accentuation has changed significantly since the RP era. Phrasing is intimately connected with grammatical structure, which tends to change less over time than pronunciation. Accentuation is also quite stable, although the stress patterns of many individual words have changed. (Lindsey 2019: 97)

However, the tunes have changed considerably in BrE as it is spoken nowadays. Among the changes in modern BrE are, for instance, a downstep to signal non-finality is perceived to be more old-fashioned and more recent models use a low pre-nuclear pitch (Lindsey 2019: 105-106). Thus, "RP made greater use of downstep than contemporary intonation" (ibid: 98). However, for the present study the H* L- pattern has been shown to be the most frequently used to indicate non-finality as well as finality. The high drop is more common in contemporary BrE. The difference between a high drop and a low drop is more of an alignment issues with the nuclear tone on a specific syllable, and it is not a difference in phonological category. This was not investigated in the present study. English now prefers a climatic high fall on the nuclear tone rather than a gradually falling pitch across an utterance, the low drop can sound authoritative (Lindsey 2019: 100). Lindsey (2019: 100) goes on that learners of English should use the more contemporary high drop for neutral statements and *wh*-questions instead of the low drop. Future research has to keep in mind the linguistic changes and the year in which the data was recorded. Even though the learners in the present study frequently used the H* L- pattern, this does not mean that it is still in use, or it aligns with different syllables and that this should be the most recent model to teach to NNSs. Lindsey (2019: 101) gives the following advice: "The simplest advice for learners interested in intonation is to acquaint yourself with the downstepped fall, but generally to be wary of using it. […] The high fall tone is more recommendable for day-to-day use". Based on varietal differences, Grabe et al. (2005: 331) suggest that learners could pick the dialect they would like to approximate to and they would have to gain long-term exposure to the dialect.

In accordance with Jenkins's (2000) Lingua Franca core, they suggest, based on their findings that learners should focus on accent placement, since this feature seems to be consistent across dialects of English (332). They also suggest that the teaching of a lot of different f0 tunes might not be as useful and that learners themselves should pick which popular patterns they like and adopt them after long-term exposure. While all prosodic features are important, one could for instance start with intonational phrasing. Some researchers do not stress the importance of intonational phrasing: "It is my hypothesis that phonetic errors of this type do not need special pedagogical treatment since they will take care of themselves as the student progresses towards intermediate and advanced learning stages" (Gutiérrez Díez 2012: 223). Gutiérrez Díez recommends that teachers should only need to approach phonological tonality errors. Yet, the concern here is how much more advanced do the learners need to be? Most of the learners in this study were already very advanced, but still seem to struggle with intonational phrasing. The importance of intonational phrases is highlighted by Gilbert (2008: 2):

> Emphasis that conveys the wrong meaning, or thought groups that either run together or break in inappropriate places, cause extra work for the listener who is trying to follow the speaker's meaning. If the burden becomes too great, the listener simply stops listening. The principle of "helping the listener to follow", therefore, is a vital one. It is so central to communication, in fact, that time spent helping students concentrate on the major rhythmic and melodic signals of English is more important than any other efforts to improve their pronunciation.

While a "corpus literacy" (Mukherjee 2002) by teachers would be desirable, simple corpus tasks could be created with the specific purpose of teaching prosody. Teachers could include stretches from corpus recordings and let learners segment and analyze intonational phrases. Through this also other prosodic features could be introduced, i.e. f0 range to signal topic initiations and to hand over turns. Mennen et al. (2014: 326) also suggest to pointing learners to cross-language differences in different parts of a tune, i.e. showing that pitch is generally higher and wider at the beginning than in later parts of a tune or how tones are used to establish a social connection with a listener in a narrative and how to create cohesion in speech. One important aspect that needs to be kept in mind when teaching "typical tones" is the following: "In reality, neither falling nor rising intonation can be said to be normal, as they served complementary functions that make each normal in particular conversational contexts. The choice of the contour depends not on grammar but on the use of the utterance" (Levis 1999: 49). The context determines which tones are appropriate and general rules cannot always be made. This again highlights the importance of teaching the pragmatics of prosody. The Prosody Pyramid proposed by Gilbert (2008) includes various exercises and suggestions for the teaching of prosody.

In addition, learners also have to be made aware of the stylistic variation of tones. While NSs are able to suppress regionally marked intonational patterns, learners are less likely to notice these and adjust them depending on the speaking style (see Ulbrich 2008). For instance, the extreme uptalk rises observed in some of the learners in the present data set are used quite inappropriately and their pragmatic intent is not always clear. Learners need to be made aware of the use of high tones and when and how to use them. Many teachers and some researchers frown upon the usage of uptalk even in native speech (see Park 2011). However, views like Park's on non-native prosody are rather the exception than the rule and many instead follow the 'learner variety' approach (Perdue 1993; Klein and Perdue 1997). In this approach, the acquisition of L2 prosody is viewed in terms of its systematicity and not in terms of their errors or deviations. Nevertheless, uptalk is an essential feature of interactive speech and needs to be topicalized and practiced. Another important aspect to keep in mind is that "[e]xpecting learners to share native speakers' perceptions of meaning is unreasonable" because "[e]ven among a group of native speakers, perceptions of intonational meaning are anything but unanimous" (Levis 1999: 42). Every NS has their own idea about what pragmatic norms are appropriate in which context (Culpeper 2018: 18). I hope that my research has shown that teachers should be encouraged to see the use of deviant prosodic features in non-native speech not as errors *per se* but instead as attempts at conveying the correct pragmatic effect of utterances, which contribute to successful conversations. While misunderstandings can arise from the use of deviant prosody, for instance, a falling tone on a question which is not marked syntactically might cause problems in comprehensibility, most of the deviant prosodic features do not impede comprehensibility but might only (if at all) elicit negative attitudes or impressions from NSs (cf. Toivanen 2007: 244). Some of the LINDSEI speakers, who stated the intent to become English language teachers in the recordings, are probably language teachers themselves by now and are models for current learners of English (see also Rosen 2019: 239). I want to focus on the root of the problem, i.e. target those instructors who teach future teachers at the university and give some advice to instructors of future generations of teachers, i.e. "what should I tell my students?" The results of this study demonstrated that the learners produced more high boundary tones, some of which were used successfully by certain students as a communicative device and interpersonal marker. Therefore, telling students to avoid using rises at the end of their utterances because they sound uncertain or anxious may not be the best solution, considering this in turn may result in a narrower f0 range. However, this is a question of quantity, as well as the tune pattern concerned. If students overdo it with uptalk instances, such as speaker GE028 who uses a rise at almost every ip break, then it is hard to focus on anything else but these rather non-native-like rises, while with speaker CZ032, these rises are more varied and competently used. Although both conversations are successful, i.e. the rises do not lead to questions, misunderstandings, or interruptions by the interviewer, the impression one gets as a listener (see also Götz 2013), is a quite negative picture of speaker GE028. Therefore, instructors should make students aware of the effect their rises have on listeners. Instructors should tell their students that listeners may have perceptual problems categorizing the rises and may not know how to interpret the pragmatic meaning of these rises and may right-away assume that the speaker is non-native and, therefore, not fully proficient in his/her L2. However, further perception experiments are needed to verify this impression. Learners use different strategies to buy more time to plan their utterances ahead, e.g. they use filled pauses and smallwords (Götz 2013), repeats (Gráf and Huang 2019), and as this study provided evidence for, they also use rises. More advanced learners may have adopted different or additional fluency-enhancing strategies (cf. Götz 2013: 109). While lower-level learners may only make use of many repeats and pauses, more advanced learners may also make use of sustained rises at the end of IUs. Therefore, more advanced learners may have changed up their strategies to buy more time. The present study has identified the following uses[83] for high boundary tones, all of which, need to be made explicit to language teachers:

- Floor holding
- New information
- Cohesion
- Interpersonal marker
- Pragmatic marker
- Comprehension checks
- Listener acknowledgement
- Uncertainty

Suggestions on how to learn/teach/asses these functions of rises as well as other prosodic features are explained in the following paragraphs.

Not only the different varieties of prosodic features in varieties of English make many teachers uncomfortable teaching prosody, but

> [p]art of this comes from confusion related to intonation terminology. Even cursory familiarity with pedagogical approaches to intonation makes it clear that the same phenomena have as many names as characters in a Russian novel. British and American pedagogical traditions to intonation use different terms and different formalisms (Levis 2005), and teachers who eclectically consult various sources cannot help but be confused. Levis (2018: 175)

Therefore, teaching materials need to be presented in a format that is accessible to learners and future teachers. Unfortunately, the material for teaching intonation in schools that is available has not much changed since the criticism Levis (1999) voiced. To be specific, there is a lack of innovation in intonation teaching materials, as well as an inadequate view of the functions of intonation, i.e. overemphasis on signaling grammatical function and conveying attitudes and emotions (see also Jenkins 2000; Derwing et al. 2013). According to Pickering (2009: 238), the attitudinal function of intonation is "something that is notoriously dependent on the context in which it appears and is thus difficult, if not impossible, to teach". Furthermore, he contends that the current materials lack a communicative purpose. His proposal involves intonation being taught in context, intonational meanings being generalizable, part of a larger communicative purpose, and connected to realistic language use.

In addition, future teachers will have to be trained on how to include prosody and how to combine this with lexico-grammatical features in their language teaching in teacher education at universities. Previous research on the prosody-pragmatics interface has found learners to make several pragmatic mistakes by choosing the same tune for various pragmatic functions, e.g. more frequent production of falling tones for all speech acts performed in a classroom setting (Ramírez Verdugo's 2008) or the overgeneralizing of rising tunes in all tag questions (Ramírez Verdugo and Romero-Trillo 2005; Puga et al. 2017, 2018). These findings by previous research suggest L1 influence and limited experience with pragmatic functions in

[83] Examples and a discussion of these uses of uptalk is provided in Section 5.5.

L2 speech and an overreliance on syntax rather than prosody (ibid.). Especially Spanish speakers seem to experience difficulty with the signaling of pragmatic functions in L2 speech through the use of intonation, because in their L1, they opt for more overt and lexicalized options possibly in combination with rising intonation in the case of tag questions (Quilis 1993: 451-452) and no intonation on pragmatic markers (Romero-Trillo 2007: 89). This is a problem, because while NSs are able to project themselves as, for instance, authoritarian or collaborative depending on the task they are involved in, NNSs fail to use their prosody to distinguish different pragmatic functions and gauge how they come across to listeners. Because if listeners assume a high linguistic proficiency, the misuse of speech acts and their pragmatic intent (i.e. wrong tune or pitch on pragmatic function) will likely to be interpreted as an unpleasant personality instead of a lack of experience with pragmatics (Tannen 1986; Pickering et al. 2012). According to Ramírez Verdugo (2008: 227), prosodic-pragmatic associations in L2 speech seem only to develop when the learners have been explicitly made aware of them in prosodic instruction. Current SLA theories propose that words are learned by being stored in the mental lexicon along with phonetic and phonological details (e.g. Pierrehumbert 2003: 149). Furthermore, not only individual words but rather lexical chunks, also called formulaic sequences, are stored and processed in the brain with their prosodic information (Pickering 2009; Lin 2012a, b, 2013). The prosody of formulaic sequences is analyzed in more detail in Lin (2018). Previous research has shown that learners of English need to acquire the pragmatic, syntactic, phonetic and phonological properties of different sentence types (cf. Colantoni et al. 2015, 2016b, Puga et al. 2017, 2018) and that learners struggle with varying their use of pragmatic devices (Götz 2013; Buysse 2017). Previous empirical work promotes the instruction of discourse markers and smallwords, which has a lot of benefits, including a possible increase in fluency, which additionally increases with longer stays abroad (Götz 2013; Götz and Mukherjee 2018; Rosen 2019). However, Rosen (2019) also showed that there are speakers in LINDSEI-GE that despite a stay abroad do not use certain smallwords frequently while also learners who have not been abroad may use smallwords just as frequently as their native counterparts. She also found considerable individual difference in the frequency of smallwords in native and GerE speech. Thus, while the focus can still be on lexico-grammatical features, the native-like prosodic patterns should be taught with it, because if the prosody is ignored, then the more fine-grained pragmatic effect is ignored and this can lead to pragmatic fossilization of these features. For instance, Pickering (2001) found in her analysis that the non-native teacher assistants did not exploit rises to establish a social connection with students and to make their speech more comprehensible compared to native teachers. The pragmatic functions signaled by rises were significantly reduced in non-native speech and Pickering recommends that more attention should be paid to the discourse meaning of intonation, because it aids the teaching of content in a classroom through an ease of comprehensibility and better relations with students. The combined analysis of fluency features and the use of uptalk rises in the present study additionally showed that learners may use these features in combination to receive more planning time, e.g. dragging the rises at the end. One suggestion for language teaching might be to teach pragmatic markers and discourse markers and their different prosodic realizations. Varying their prosodic realizations on pragmatic markers and avoiding the excessive use of high boundary tones may help learners to be perceived in a more positive light. This is not to say that learners should not use high boundary tones at all, but they should be made aware of their pragmatic effects and the attitudes they trigger by native listeners. As Levis (2018: 181) rightly points out, “[u]nlike vowels and consonants, intonation is unlikely to be heard as right or wrong, and unexpected use of intonation is much more likely to be judged not as a pronunciation error but a social failing”, which can lead to “misjudgements of the speaker’s intent” (Gumperz 1982: 132). It can also trigger attitudes such as uninterested, morose, or aggressive (Pascoe 1996: 114). Previous research has shown that it is possible to reduce a foreign accent through the teaching of pitch variation (e.g. Hincks 2004; Levis and Pickering 2004; Hincks and Edlund 2009). Nevertheless, not much is known about the pragmatics of prosody and more research is needed for the development of teaching materials and teacher training (cf. Riesco-Bernier 2012). The acquisition of tunes for certain pragmatic functions would improve with more focused instruction. The observations made in the present study have several implications for research into and the teaching of the pragmatics of prosody. One implication would be for the unit of investigation to contain smaller units such as ips and their pragmatic intent, since larger units can have several pragmatic functions, even within smaller units it is sometimes difficult to decide which function a speaker intended to use. Therefore, narrowing down the unit of analysis definitely helps to analyze prosody-pragmatic connections (see Staples 2015; Fernández and Staples 2021).

The results of the present study also have implications for language assessment, since several characteristics of spoken interactions often employed in language testing scenarios were found. Specifically, I would like to discuss and add suggestions for more fine-grained CEFR descriptors to rather vague mentions of prosody in the CEFR companion volume (Council of Europe 2018: 136). Most of the learners in the present study can be described as having a C1-ability (at least in LINDSEI-CZ and -GE), although a few learners with C2 and possibly B1 levels were included as well. A look at the new CEFR descriptors for prosodic features in the phonological control grid shows that the pragmatics of prosody takes center-stage for all CEFR levels, with “strong” (B1), “some “(B2), “occasional lapses” (C1) to no L1 influence (C2) being expected on the prosodic realizations of pragmatic functions, i.e. to convey a message intelligibly and effectively through the use of prosody (i.e. stress, intonation, and rhythm). Therefore, I will focus on the prosodic realizations of pragmatic functions, i.e. whether learners are able to “express precisely what he/she means to say” (C1 ability) and their “finer shades of meaning” (C2 ability) and the degree of L1 influence for the B1-C2 levels. More fine-grained descriptors are needed, because for language teachers it is rather imprecise what is meant by these descriptors and how exactly one would assess them. Another question to be asked is whether more precise prosodic features such as f0 range (cf. Cauvin 2013) should be included in the phonological control grid of the CEFR descriptors as well. The present study has shown that the prosodic productions of learners are conditioned by a multitude of factors. Particularly interesting are the effects of the learners’ fluency, proficiency level (CEFR levels), and exposure to the TL (i.e. stay abroad) on the prosodic variables analyzed. These factors need to be looked at more closely, since if the learners’ L2 prosody is indicative of a certain proficiency level, or fluency use, or the length of a stay abroad, this information could inform the CEFR descriptor scales that are based on empirical evidence. Taken together, the present study showed that the CEFR proficiency levels of phonological control are not ideal indicators of the learners’ prosodic performance, i.e. the CEFR levels do not correlate strongly with any of the prosodic measures or other indicators of advancedness. With the exception of speech rate (moderate correlation), none of the correlations showed a strong effect of CEFR levels on the prosodic variables investigated. Since there is no strong correlation between CEFR levels and any of the analyzed prosodic features, it can be concluded that they can be regarded as separate measures of prosodic fluency and they should be used in language assessment and teaching, since these prosodic deviances seem to persist even in advanced learner language. For the prosodic variables and their correlation with CEFR levels further research is needed with larger learner groups and more balanced information on extralinguistic factors to validate this tentative suggestion. However, since no clear-cut results were found, it may be difficult to include prosodic behavior as a variable in the CEFR descriptor scales, especially because some of the prosodic features have been shown to differ by L1 and speaker sex. Overall, since most of the learners exhibited similar tendencies in their L2 prosodic productions, as summarized in the global discussion in Section 5.1, L1 influence on L2 prosody is rather minimal. Therefore, it may make sense to include descriptors in the CEFR that are common to all learners, i.e. narrower f0 span for HL-tunes, wider and higher f0 for HH- and LH-tunes, more high edge-tones, and increased uptalk usage. More L1-based differences were found for intonational phrasing. Therefore, based on the prosodic level that is assessed, descriptors that all learners have in common (i.e. developmental features) and L1-based features have to be distinguished and assessed according to speakers' L1. Thus, for f0 range and tonal choice, it makes sense to assess the learners on the same scale, while for intonational phrasing there may be both proficiency, gender-, and L1-based differences, which may make it more difficult to assess learners on the same scale. As far as f0 range is concerned, I am not entirely convinced by Cauvin's (2013: 85) suggestions on rating learners' f0 range based on specific STs, even though the overall idea of rating learners' f0 range in exclamatives is not wrong *per se*, i.e. lower than 6 STs as a fail, 9 STs as a pass, and 12 STs with a distinction grade. Even though I would agree with this general trend, this is, however, difficult to measure by language proficiency tests and for teachers who grade students. F0 range is highly context-dependent and Cauvin acknowledges that the sentence type, position-sensitivity and the pragmatic intent are not taken into account. The present study has shown that learners produce a narrower f0 span for HL-tunes and no significant difference in f0 level for HL-tunes were found. However, when it comes to HH- and LH-tunes, which are usually produced at the periphery of IPs for mainly continuative or checking purposes, they are produced with a significantly higher and wider f0. In the case of the CzE and GerE speakers, this can be explained by a more excessive use of the checking function. However, in the case of low tunes (finality and handing over turn) NSs produce significantly lower f0 levels. Thus, in sum, learners overhit the f0 in high-tunes and underhit f0 in HL- and low-tunes. This is complicated further by the different strategies that NSs employ to signal their pragmatic intent, i.e. AmEO speakers were shown to use a high f0 level and a wider f0 span to signal emphasis, while the NBrE speakers only used a higher f0 level and not span. While it is recommended to use NS data to develop language tests (Barker 2010: 635), another problem that has to be mentioned at this point is that monolingual speakers of English were chosen for the present study, and it has to be taken into account that bilingual speakers of German/English produce intermediate f0 values as compared to monolingual speakers of both German and English for different speaking styles (cf. Scharf-Rethfeldt et al. 2008: 127). Future studies should include bilingual speakers of English and these results should be taken as a benchmark instead of the monolingual NSs. Furthermore, not only learners' linguistic performance of prosodic features but also their interactive skills and how effectively they communicate should be assessed. For instance, one factor that needs to be taken into consideration is the level of accommodation. Accommodation of f0 range could be seen as one advanced factor in language assessment. Interestingly, the MurE female speakers were the only speakers who were shown to entrain their f0 level the least to that of their interlocutors. While in the case of the MurE speakers this may be an indication of a lower proficiency level, in the case of the CzE and GerE speakers, the rater (if involved in an interactive task with the learners) should ask him-/herself whether they are speaking with an exaggerated teacher intonation (as discussed in Section 5.5). Because if learners accommodate, this is generally a sign that they are able to perceive and produce native-like prosody. Furthermore, the use of uptalk and checking in general may be an indication of a higher proficiency level due to the fact that learners are trying to make their pragmatic intent of their utterances overly explicit to the listener. Learners may overly rely on the use of intonation to fulfill functions that NSs use pragmatic markers for (see discussion in Section 5.5.). Learners could be assessed on the following criteria that could be seen as consistent indicators of proficient and native-like prosodic productions across learners from different L1 backgrounds:

- Learners who use wider f0 span in HL-tunes in declaratives
- Learners who use lower f0 level in HH- and LH-tunes in declaratives
- Learners who use fewer instances of uptalk, checking, and insecurity, and possibly use more discourse/pragmatic markers instead
- Learners who accommodate to their interlocutors in f0 range (pay attention to own intonation, i.e. the raters/testers intonation)
- Learners who produce long and coherent IUs without an excessive use of rising tones, connectives, hesitation, filled and silent pauses, and elongations

Since fluency is already mentioned elsewhere in the CEFR, it will not be listed here. Even though it is measured on prosodic units in the present study, the overall trend for learners to exhibit a slower speech rate found in previous research was corroborated. In sum, even though I agree with the overall idea to focus on the pragmatic intent and their prosodic realizations in the CEFR descriptors, they are in need of a more fine-grained level of including f0 range, prosodic accommodation, as well as intonational phrasing, instead of just listing stress, intonation, and rhythm as examples of prosodic features. Additionally, depending on the prosodic feature that is assessed, different scales for not only L1 background but also gender have to be distinguished. I admit that it is extremely difficult to distinguish, perceive, and assess these prosodic features and more perceptual experiments are necessary to add more empirical evidence to the prosodic features that raters and language teachers should assess. An aspect the present study did not investigate empirically is the perception of L2 or L1 prosody by native or non-native listeners. There are many open questions about how prosody in general is perceived, and how it contributes to intelligibility and comprehensibility. According to Levis (2018: 151), three features of intonation are particularly important for intelligibility, i.e. relative prominence, tune, and f0 range. These features rarely impact intelligibility at the word level (as describe by Jenkins in the lingua franca core concept) but rather at the level of the message interpretation (171). One of the core arguments Levis (2018: 178) makes is that we have to move away from a word level analysis and go beyond that and then the effect of intonation on intelligibility will become clear. Future perception experiments could investigate how longer vs. shorter rises are perceived in relation to the pragmatic function they fulfill, i.e. the perception of rise meaning based on duration (Armstrong et al. 2015). A fundamental issue in perception research was also proposed by Ward and Gallardo (2017: 24), who asked the question of which prosodic differences actually matter in perception and which interpersonal and social factors contribute to perceiving differences. Previous research found a correlation between the learners' ability to delimit IPs and their rated oral proficiency in the L2 (Wennerstrom 1998; Kang et al. 2010: 556). According to Götz (2013), intonation is one of the perceptive fluencemes that contributes to an impression of an overall fluency on the side of the listener. In her analysis of NS ratings, she found that intonation was perceived to impede fluency of individual speakers. Derwing and Rossiter (2003: 11) received far more negative comments on L2 learners' prosody from spontaneous narratives elicited from judges than on phonological or morphological errors. Spezzini (2004) also found that NS judges were mostly influenced by prosody and the overall fluency of speakers when rating comprehensibility in an L2. Gut (2009) found significant correlations for the mean length of syllables with full, reduced, and deleted vowels and the length of phrases. Future studies will have to investigate the effect of intonational phrasing directly. Mennen et al. (2012: 2258) findings suggest that listeners may not be evenly influenced by different parts of an IP and that they may base their judgments on the beginning of an IP in L1 productions. Götz's (2013) findings seem to suggest that the perception is mainly based on the end of utterance intonation, i.e. uptalk in L2 productions. It still needs to be determined which parts of an IP influence a listener's judgment of f0 range or tonal choice.

> The perceived impression that a spoken utterance is not native is due to the culmination of a combination of segmental and suprasegmental phonological and phonetic characteristics entrenched in the speakers [sic] native language (L1) or it results from different types of interplay between those characteristics of the L1 and the second (or foreign) language (L2). (Ulbrich 2015: 1)

In sum, these studies on the perception and correlation of prosody in combination with other linguistic features show that prosody has a big impact on attitudes towards and perception of fluency by native raters. These findings have implications for language assessment, CEFR rating, and language teaching, because they can provide further empirical evidence for which prosodic features are perceived in what way by individuals who could possibly rate the prosodic productions of learners, i.e. teachers. Overall, at C1-C2 levels, raters should mostly pay attention to the use of f0 and whether the pragmatic intent is clear, which should be virtually free of L1 influence. For B1-B2 levels, the focus should be the same as for C1-C2 levels but raters can be more lenient/accepting of L1 transfer phenomena. Based on the results of the present study, general trends were observed and specific suggestions for and questions to be asked during language assessment without using speech analysis software can be proposed:

- Using rises at the end of IUs does not make learners non-native-like *per se*. The rater/instructor has to differentiate between competently used rises and more foreign-sounding ones on the basis of their realization, distribution, and functions these rises may fulfill. Just as in NE speech, uptalk in non-native speech may be distributed differently and used for different pragmatic functions by men and women. Female NNSs of English may be organizing their discourse in a more complex way, just like NE females (cf. Levon 2016). Thus, if different pragmatic functions of rises can be identified by the rater, a higher proficiency level of the learner could be assumed because speech might be fluent enough and lexical retrieval might be more automatic. The learners may, therefore, afford to pay more attention to their f0 patterns and involve the interlocutor into the conversation by structuring their discourse through the use of uptalk. More attention should be paid to rises than falls, since rises are acquired later in the SLA process (cf. Grosser 1997; Gut 2009). Questions raters could ask during assessment (possibly in form of a questionnaire) could be the following:

- Do the learners repeat the same type (sustained, rising, etc.) and exaggerated realizations (extremely high or low f0) of tones to signal certain pragmatic functions? For instance, if learners are involved in a monologue, do they vary their use of tones to signal continuation and to elicit feedback from the interviewer or do they use the same type and realization of tones?
- Do they use extremely low f0 for to signal a handing over of turn in dialogues and/or do they use extremely high f0 to signal continuation and checking in answers or monologic tasks?
- Here the rater can pay attention to his/her own behavior. Does the learner elicit back-channeling feedback and/or even another question/statement from me? Or do I generally back-channel frequently and may therefore elicit more rises/uptalk from the learner? Am I speaking in an exaggerated teacher talk style, which may result in accommodation by the interviewee to my own speech? Are my own prosodic productions within the learner's phonetic repertoire based on gender-pairing and English dialect spoken/aspired to?
- Does the learner use pragmatic markers as well or overly rely on intonation to signal pragmatic intent?

- Intonational phrasing also contributes to effective and intelligible communication and an excessive use of fillers and tones may make it hard to process and comprehend utterances. A conjunction of utterance-binding and/or separating features together with excessive rises or falls can lead to non-native intonational phrasing patterns. While on the C1 level occasional lapses in the control or prosodic features are allowed, on the C2 level they should not occur. CzE female speakers can be expected to produce fewer IU breaks and GerE female speakers to produce more IPs, which are possible indicators of an L1 influence for both groups. If these tendencies disrupt a smooth and intelligible spoken discourse and the rater has difficulties to decide when to take over a turn or when to ask a question then the learner would have to be rated more negatively on his/her performance.

Furthermore, I want to point towards ideas on how to train future teachers to learn and teach prosody themselves: Hyde (1994) suggests letting teachers in training analyze their own speech, because "if authentic data is used it is impossible to analyse the data fully, as only educated guesses can be made as to speakers' pragmatic intentions and psychological outlooks. […] they will have first-hand knowledge of speaker intention and the choice of tone and key can be analysed from an insider perspective" (Hyde 1994: 147). Another teaching method is proposed by Pickering (2004), who suggests theater exercises to increase confidence and to give non-native teacher assistants the opportunity to explore their voice range. Besides using recordings and mainly training the auditory channel, also different channels could be used to teach prosody. Acton et al. (2013), for instance, encourages the use of a haptic method to teach intonation, similar to a total physical response approach, where learners use movement and touch to coordinate their body with both segmental and prosodic features. Thus, in sum, the following suggestions for teaching future teachers can be made based on the findings in the present study:

- First instructors should discuss with students how prosody can be defined and analyzed. For consistency, the ToBI annotation system should be used instead of having the students coming up with their own prosodic symbols, which would just contribute to the confusing surrounding prosodic annotation, analysis, as well as terminology. Not just the prosodic symbols but AM terminology could be used so that the future teachers in teacher training can refer to these terms as well and that way contribute to a more common understanding of what prosodic features are and that they can be referred to in class by students.
- Instructor should focus first on intonational phrasing and the effect of f0 as a cohesive device, discourse management, as well as relationship-building device.
- Based on the learners' L1, the instructor has to make the learners aware of utterance-binding and –separating prosodic and fluency phenomena. This again highlights the interconnectedness of pragmatic, fluency, and prosodic knowledge. For instance, if CzE learners used more discourse markers, and fewer utterance-final elongations and filled pauses, as a result their intonational phrasing may become more native-like. GerE learners have to be made aware of the use of sustained/extreme rises in conjunction with silent pauses, which may lead to the overproduction of IUs. SpE speakers did not show any significant deviances from the NSs when it comes to intonational phrasing, although previous research has also shown that SpE speakers transfer fillers from their L1 to the L2 and, therefore, may produce non-native-like IU frequencies.
- Teach the prosody of pragmatic markers in conjunction with fluency phenomena. Sustained rises can be used as fluency-enhancing strategies. For C2-levels, instructors should focus on the finer shades of meaning and how prosody can be used to achieve this goal. Especially tones and tunes should be analyzed by instructors and students and how they can be employed to signal different pragmatic functions such as floor holding, signaling new information, showing cohesion and disjuncture, comprehension checks, etc. If the focus is on the pragmatics of prosody, then this may improve the other intonational dimensions at the same time. For instance, if learners know how to use rises appropriately to

signal cohesion, instead of using extreme rises that come across as signaling uncertainty, then their f0 range will be reduced for these tunes as well.

- Instructors should make students aware of the appropriateness of use of uptalk. In which situations would they be beneficial (e.g. in the classroom see Pickering 2001) and where they would not be appropriate (e.g. job interview).
- Instructors could have students (future teachers in teacher training) prepare small monologues and have other students listen and note down prosodic features that "stand out". Instructors should make their students aware of their own speech and they should be able to reflect on their use of rises and falls as communicative devices. Best would be to involve the students in an interview and play the recordings back to them in order to make them aware of how they used rises and falls (and overall prosodic patterns) in conversation themselves.
- Instructors could have their students collect recordings of a model speaker of an English variety that they find pleasant and/or aspire to speak who is involved in a dialogue. With the help of speech analysis software (e.g. *Praat* or *Elan*), the students could analyze the f0 line and intonational phrasing and find instances of each pragmatic function and note down their prosodic realizations (e.g. f0 range used for certain functions). Then they could try to imitate some of the utterances made by the model speaker and compare them side-by-side with their own productions. That way the instructors would also have a collection of pragmatic functions along with an analysis of the prosodic parameters and could extend this collection for future student groups. Ideal would be to have a public platform for teachers available with such data. Of course, if corpus data is used, these cannot be made publicly available.
- Instructors, especially if they are non-native, should become attuned to different varieties of English and regional differences in terms of prosody (e.g. AmE vs. BrE) and give students advice on how to become familiar with the prosodic patterns of a certain variety of English. Future teachers would have to be trained for this in special teacher training sessions.

Additionally, Levis and Sonsaat (2019) provide guidelines for quality teacher education for pronunciation teaching, including which attitudes, values, knowledge, and skills teachers should have when teaching pronunciation.

Not only do the results presented in this study have implications for prosodic and pragmatic theory, teacher training, and language assessment, but also for speech technology. For instance, the expression of uncertainty in non-native speech is often used in interactions with online tutors (cf. Wollermann 2012: 96). Previous research has shown that if the system adapts to the uncertainty of the user, this has a positive effect on the learning gains of students (Pon-Barry et al. 2006). Not only can uncertainty be identified by computers but computational and human-perception studies have found that emotions such as anger, happiness, and sadness can be identified by human listeners and automatically classified using acoustic and prosodic features with fair to good accuracy (Ang et al. 2002; Yuan et al. 2002; Park and Sim 2003; Litman and Forbes 2003; Bitouk et al. 2010; Dimos et al. 2015). Interestingly, previous research has found a classification accuracy of 79-91% of a speakers personality based on their prosodic stylization (e.g. Reichel 2015). I am certain that the developments in speech technology and prosodic research will produce software programs that will help learners to improve their prosody. Speech analysis programs that give immediate feedback on how one delivered a message or presentation, e.g. which prosodic tunes create meaningful contrasts between an L1 and L2, how fluent a speaker is, were pauses used to let words sink in, can enthusiasm be detected in the voice, etc. (Chun 1998: 81; Hincks 2004: 64). This will not only be beneficial for language learners but also for NSs who would like to become more effective public speakers (Hincks 2004: 65; Hoque et al. 2013). Initial attempts at feedback for intonation contours (f0 range) have been made by Hincks (2004). Chun (1998: 81-82) suggested that software programs should be able to distinguish meaningful intonational features with regard to the direction, range, speed, and place of pitch change and would have to go beyond the sentence level to address the various functions of intonation, i.e. grammatical, attitudinal, discourse, and sociolinguistic. Additionally, Chun (1998: 81) suggests that learners should start analyzing utterances with a minimum of voiceless sounds to make the deciphering of f0 curves easier for them. I also suggest using recordings with a great audio quality, ideally carefully read speech for first encounters with speech analysis software.

5.6.4 Summary

First, Section 5.6.1 revised the different SLA predictors analyzed in the present study and identified and discussed further predictors that might be used in future research. Second, the intonational dimensions of the LILt model were discussed in the light of the results of the present study and a multifactorial model of non-native prosody was presented and discussed to explain the results. Third, implications on different aspects of language teaching and speech technology were discussed and suggestions for language assessment were made. The following section summarizes the major findings of the present study and provides an outlook on the field and suggestions for future research.

6 Conclusion and outlook

The last two sections summarize the major findings of the present thesis, highlight new findings, give suggestions for future research, and provide an outlook on the topic.

6.1 Summary of major findings

One of the central aims of the thesis was to compare the L2 production in two speaking styles with those in native speech. The overriding aim of the present study was to contribute to a better understanding of not only how learners differ from natively produced prosody but also if the prosodic features are shared by these distinct L1 varieties. This dissertation has given an account of declarative intonation patterns across seven different varieties of English, i.e. AmEO, NBrE, SAmE, CzE, GerE, MurE, and MadE. The data in the present study was analyzed pragmatically, phonologically, and phonetically by quantifying the occurrence of ips with certain pragmatic functions and their tonal choice and f0 range. Table 52 provides a summary of the NNS results and in how far they deviate from the NBrE benchmark. The overview demonstrates that all learners deviate from the NS norm into the same direction but L1-based differences were observed for the strength of deviance for each prosodic feature. The deviances in Table 52 are numbered from the most deviant (1) to the least deviant (4), which are however still significantly different from the Ns norm. If learner groups showed equal values for a prosodic feature, they received the same number. The gray cells in Table 52 highlight the prosodic features in which the NNSs show opposite trends of deviances from the NS norm and from each other. Empty cells demonstrate prosodic features where certain learner groups did not significantly deviate from the NS norm. In total there were 17 individual prosodic features included in the overview and a summaries of all deviant and most deviant features are provided at the bottom of the table. To answer the original four research questions, the results of the present study seem to suggest that a non-native interlanguage intonation system exists for the majority of learners from all three L1 backgrounds, which corroborates previous research (e.g. Ramírez Verdugo 2006b). It can be said now that spontaneous speech produced by all NNSs is characterized by slower IUs (measured in WPS), a narrower f0 span in HL-tunes, a wider and higher f0 range in high-register tunes (LH- and HH-tunes), and more high boundary tones and high-register tunes. Moreover, the learners produce a native-like f0 level in HL-tunes as well as a native-like tone frequency. Additionally, all learners show entrainment effects, and all learners use more pragmatically diverse prosody (i.e. emphasis, insecurity, checking, etc.) (RQs1-3a). Applying Mennen's theory to the case of learners from three different L1 groups, the analysis revealed that the learners differ on the distributional, realizational, and functional dimension of intonation from native speech (RQ3a). Differences between learner groups were also identified (RQ1): SpE speakers frequently produced high boundary tones, but not as many as the CzE and GerE speakers. GerE and CzE speech was marked by a high frequency of uptalk, for which no significant phonological differences were found between the two learner languages. From a phonetic perspective, CzE and GerE produced extremely similar f0 in the uptalk tunes (except for two variables). For IU frequency gender- and learner variety specific effects were observed: The GerE female speakers were the only ones that produced significantly more IPs. CzE female speakers produced fewer IU breaks. SpE speakers produced native-like IU frequencies but produced the narrowest f0 span for HL-tunes and widest for LH- and HH-tunes, and tended to entrain their f0 less to their interlocutors. MurE speakers produced a higher f0 level and wider f0 span for LL-tunes and the majority of high pitch accents. The CzE and MadE speakers produced the slowest IUs (WPS). Thus, while the learners all show similar trends in their prosodic productions, they differ from each other mainly on the distributional and realizational level. Some distributional and realizational differences were attributed to gender, interviewer role and gender, speaking style, fluency variables, proficiency, and position-sensitive differences (i.e. position of tones within an IU) (RQ3b). The overview in Table 53 shows that gender and fluency seem to be the most influential factors for most of the prosodic variables, while also minor age, speaking style, stay abroad, and proficiency effects were found. Overall, the learners show many structural and functional similarities but the overview in Table 52 shows that MurE speakers produce the most non-native-like prosody, while the GerE prosody is the most native-like. The table also shows that CzE and GerE speakers have many features in common, while MadE and MurE speakers also share many features. While the GerE speakers are overall closest to native-like patterns (which further depends on which native variety they are compared to), they are closer in the realizational dimension (i.e. overall f0 range) than in the distributional one (i.e. IUs, and tones and tunes). For the SpE speakers (MadE speakers more so than MurE speakers) it is the exact opposite; while they are closer to native productions on the distributional dimension (i.e. IUs, and tones and tunes), they are furthest away from the realizational one (i.e. f0 range). The repeated uptalking feature in non-native speech is probably one of the features that sticks out the most, also in the opinion of native raters (e.g. Götz 2013) which marks them as foreign (RQ4). In sum, the problem with foreign-sounding uptalk rises is a distributional one (more frequent), a realizational one (hardly any variation in the shape and f0 range of the rise and overall tune), and a functional one (due to the same shape and tonal parallelism, the same pragmatic function is repeated by most speakers).

Table 52. Summary of all non-native-like features by interlanguage

	Prosodic features	CzE	GerE	MadE	MurE
IUs	**IP frequency**	1) fewer IPs	1) more IPs	---	---
	ip frequency	1) fewer ips	---	---	---
	IP speech rate	2) slower	3) slower	1) slowest	3) slower
	ip speech rate	1) slowest	3) slower	1) slowest	4) slower
Tones	**Tone frequency**	---	---	---	---
	Edge tone frequency	2) more high edge tones	1) the most high edge tones	3) more high edge tones	4) more high edge tones
	Pitch accent frequency	2) more low pitch accents	1) the most low pitch accents	3) more low pitch accents	1) more high pitch accents
Frequency of pragmatic functions		1) vary the most	2) vary more often	3) vary slightly more	4) vary slightly more
F0 range	***F0 span* for HL-tunes**	3) narrower	2) narrower	1) narrowest	1) narrowest
	***F0 span* for LH-tunes**	3) wider	2) wider	1) wider	1) wider
	***F0 span* for HH-tunes**	4) wider	3) wider	2) wider	1) widest
	***F0 span* for LL-tunes**	---	---	---	1) widest
	***F0 level* for HL-tunes**	---	---	---	---
	***F0 level* for LH-tunes**	2) higher	2) higher	2) higher	1) highest
	***F0 level* for HH-tunes**	1) higher	2) higher	1) higher	2) higher
	***F0 level* for LL-tunes**	---	---	---	1) highest
	Entrainment	---	---	1) entrain less	1) entrain less
Total of deviant features		12/17 (71%)	11/17 (65%)	11/17 (65%)	13/17 (77%)
Total of most deviant features		5/17 (29%)	3/17 (18%)	6/17 (35%)	8/17 (47%)

All in all, this study has gone some way towards enhancing our understanding of prosody and meaning in native and non-native speech. From the SLA perspective taken in Section 5.6, the results of the present study mostly fell within the predictions of several current SLA theories, i.e. the SLM, BV, and LILt. However, contrary to the predictions of the BV theory, learners used prosody to fulfill a variety of pragmatic functions, to mark transition to reported/quoted speech, and to entrain to native speech.

Table 53 provides an overview of the SLA-predictors and their effect on the prosodic variables. The overview does not demonstrate the directionality of the effects. The direction of the effects of the most important SLA-predictors on the prosodic variables were discussed in Section 5.6.2 with the help of an explanatory model. The overview includes the results of the various statistical models with NSs-NNSs contrasts (medium gray cells) as well as NNSs-NNSs analyses (dark gray cells). The light gray cells highlight extralinguistic factors that were only analyzed in the statistical models including NNS-NNS contrasts.

Table 53. Summary of SLA-predictors and their effect on the prosodic variables.

	AGE	SEX	STAB	CEFR	SPEAKING STYLE	FLUENCY	INT_SEX	NO_OFL	YOEAU	YOEAS
Tone freq.	X	X			X	X				
High edge tones		X	X		X		X	X		X
IP freq.		X			X	X				
IP speech rate				X	X	X		X		
ip freq.		X				X			X	
ip speech rate	X			X		X				
F0 level		X			X	X				
F0 span		X	X		X	X				
Uptalk freq.	X	X	X		X					

The findings of this study that corroborate previous research are listed below:

- Universal patterns in the acquisition of L2 prosody exist for L2 learners from different L1 backgrounds, since learners make similar 'errors'
- F0 span is narrower in non-native speech for HL-tunes,
- F0 level for HL-tunes is similar in native and non-native speech,
- There are position-sensitive f0 range differences,
- High boundary tones and LH-tunes are used more frequently in non-native speech,
- Learners produce a slower speech rate.

The current study was novel in the sense that it compared the speech of learners from three different L1 backgrounds within the assumptions of the LILt framework combining prosodic-pragmatic analysis with an array of prosodic features. These are the new findings:

- There are differences on the distributional, realizational, and functional intonation dimension into the same direction for all three learner varieties, which are mainly due to developmental issues and some L1 transfer effects. The degree of deviance on each intonational dimension depends on the L1, L2 proficiency, and gender,
- There are gendered differences on almost all prosodic variables, particularly in intonational phrasing behavior gender-based differences were found (besides f0 range),
- There are significant tune-based differences in f0 range: All learners exhibit problems with the realization and distribution of high-register tunes (besides producing a narrower f0 span for HL-tunes). The high-register tunes (LH- and HH-tunes) of all learners are more extreme, i.e. higher in f0 level and wider in f0 span:
- Form meaning relationship: Socio-pragmatic functions of high boundary tones and LH-tunes used by NNSs to signal uncertainty, cohesion, and used as an interpersonal marker and pragmatic marker. Therefore, high boundary tones can be viewed as an indicator of the proficiency level of learners and these were suggested to be examples of prosodic drift and part of an prosodic exploration phase in the SLA-process. Avoidance strategies have been proposed as well.

On a methodological level, the present study has illustrated the advantages and limitations of a multifactorial corpus-based approach to studying the effect of several extralinguistic factors on prosodic improvement. Despite the drawbacks of the corpus selected for the present study (see discussion in Section 3.8), the corpus offered valuable insights into L2 prosodic production and acquisition, which may be reproduced in future research with other corpora or more detailed analyses with the present corpus. The present study has also shown that it is possible to combine quantitative and qualitative analysis with such a corpus, while the quantitative analysis has shown that the uptalk feature is indicative of non-native speech, which applies to all three learner varieties in the present study. Consequently, this has lead most intonation researchers to conclude that this is correlated with a higher level of insecurity on the part of the learners. However, the quantitative analysis has shown that uptalk is also used for a variety of pragmatic functions that may enhance communicative success, thus, demonstrating that a quantitative analysis alone does not paint a complete picture of the use of high edge tones by NNSs. Further methodological contributions the study has made are listed below:

- It is worth investigating rises in more detail with more detailed pragmatic functions (see also Staples and Fernández 2021), because not all rises in learner language are the same. The development of IPrA might make more detailed analyses of rises possible,
- It depends which native variety learners are compared to. SAmE speakers are the most deviant from NBrE prosody and show more similarities to learner language in some intonational phrasing parameters and the distribution of high-register tunes,
- Combining intonational phrasing measures with fluency measures (not a new approach) is a rewarding endeavor,
- The present study points towards difficulty of measuring prosodic improvement with one proficiency indicator alone (multiple indictors might be more revealing),
- Combining several prosodic features gives a more complete picture on the deviations along intonational dimensions.

Comparing the three learner groups with each other, it can now be concluded that they all display a systematic development toward their target. The pedagogically relevant results were already discussed in Section 5.6.3 but overall, the results show that a learners' L1 background seems negligible, and that attention should be paid more to the production of high-tones and tunes (realizational and distributional dimension) and their pragmatic effects in ELT, which seems to be a developmental factor that all learners have in common. The results support the assumption of rises developing more gradually and being more difficult to produce (Ohala and Ewan 1973; Crystal 1986; Grosser 1997; Snow and Balog 2002; Mennen et al. 2010). The present study also showed that the acquisition of prosody is a multilevel process that depends on the attainment of other prosodic features, i.e. intonational phrasing development influences the production of tones and tunes and, therefore, needs to be addressed in language teaching first (for some L1 groups). Theoretically, the present study reinforces the need of a further development of L2 prosodic acquisition theories and models that incorporate predictions based on the multilayered nature of L2 prosodic acquisition, factoring in the learners' proficiency levels and native target model they wish to approximate to.

6.2 Outlook

While having answered some questions about prosody and meaning, the research study has also given rise to additional questions in need of further investigation. Throughout the discussion, suggestions for further studies and experiments were proposed. The following ideas are by no means exhaustive but just include a few interesting directions that could be taken up with the present data set or with new data. The findings of the present study suggest the following opportunities for future research:

- While the present study has focused on rises and how they differ by pragmatic function, future research could also investigate the use of falls as cohesive devices (cf. Wichmann 2000). For instance, Wichmann (2000: 85) demonstrates with examples from the SEC corpus how syntactic and prosodic features are used in combination, and that falls at phrase boundaries show both conceptual breaks and cohesion. Equally, there are still many unresolved questions related to high tones. Future work needs to establish how rises are used in native and non-native speech to create cohesion, to ask questions, and especially how they are used to establish relations with an interlocutor.

- The present study has identified various pragmatic meanings of high boundary tones. However, no further distinction was made between more fine-grained distinctions of pragmatic functions of uptalk and list intonation. Any type of listing of items was annotated as "listing". In future studies a distinction can be made between different types of list intonation (see Burdin and Tyler 2018: 99-100 for a summary of different types). Burdin and Tyler (2018) found that list intonation patterns in AmE depend on the "belief states of the speaker" about a "listener's knowledge state", i.e. if the speaker believes that the listener knows the items to be listed, and then plateau intonation contours are used. However, if the speaker believes that the items to be listed are unknown, rising contours are used. Thus, plateaus are used to "remind" and rises to "inform". Speakers in the present data set used uptalk in list intonation to check if the interlocutor understands, but more fine-grained analysis would have to be undertaken with the data in the present study or new data to see whether such differences apply to NNSs.
- Previous research has shown that interviewers may change their conversational behavior depending on their dialogue partners, e.g. in the duration of overlaps and the frequency of backchannels and prosodic features (e.g. Weingartová et al. 2014b; Smith 2007). Levitan (2014: 26-27) warns that comparing the speech of conversational partners with speech in separate conversations may be flawed because findings may be rather attributed to the "circumstances independent of the dynamic between interlocutors". For instance, if ambient noise increases, both speakers may speak louder at the same time, or when both interlocutors speak in their L2, they may speak more slowly independently of each other. With the help of the extended LINDSEI-CZ L1 recordings[84] (n=31), one could analyze speaking style differences in more detail. Additionally, future studies could analyze the speech of the interviewers in LOCNEC and LINDSEI, and see whether these speakers adhere to a personal prosodic style, which is consistent across interactional and gender pairings with different conversational partners or whether they entrain to the speech of the specific learners. Previous research has found that speakers diverge from each other when speaking to the opposite sex (and when attracted to each other) (Michalsky and Schoormann 2018). Future research can examine the speech of the same individuals in both their L1 and L2 with different conversational partners and exploit power relations between speakers (frequency code: Ohala 1984).
- Further experimental investigations are needed to determine learner attitudes towards varieties of English. In particular attitude research coupled with learners' actual productions and ability to perceive different varieties of English based on their prosody could have promising insights into ultimate attainment and the relation to reference accents. Future research might want to establish whether learners are actually able to produce target-like L2 patterns they claim to attain to. In addition,

[84] These are not publicly available.

certain intonation patterns may be avoided on purpose because learners do not want to sound 'too British'.[85]

- Learners often use different types of movement, eye contact, and gestures when speaking in the L2. Future research should incorporate an analysis of video material if available. Especially prosody has been said to correlate with eyebrow movements and hand movements (Cavé et al. 1996; Wollermann 2012). It would be interesting to see whether learners from different L1 background (e.g. Romance languages vs. Germanic vs. Slavic, etc.) make use of such channels and where the differences lie.
- More comprehensive analyses of sociophonetic variation and change are needed (Nance et al. 2018). Especially more updated NS recordings are needed, because prosodic features change and the 1995 LOCNEC models may not be an appropriate prosodic model anymore (see also Lindsey 2019). The extended LOCNEC (Pérez-Paredes 2010; Pérez-Paredes and Bueno-Alastuey 2019), which includes 28 extra interviews from the British component from Manchester Metropolitan University (Aguado-Jiménez et al. 2012), could be exploited for more regional prosodic variation.
- As the overview in Jarvis (2000: 260-261) shows, there are many factors that influence L2 production, e.g. age, personality, cultural background, etc. Many of these factors are difficult to measure and are not always feasible; however, these features should, if possible, be included in future collections of learner data in order to offer a holistic picture of learner language and the effect of extralinguistic factors.
- The final suggestion to be made leans towards Bolinger (1986: Preface), who stated that "intonation is too important a subject to be left just to linguists". Thus, on a wider level, research on prosody at various intersections of linguistics and other fields such as psychology, sociology, robotics, etc. is needed.

The past decade has seen a renewed importance in L2 prosody. The next decade is likely to witness a considerable rise in L2 prosodic studies and will inevitably lead to advancements in language teaching and learning. In addition, recent developments in statistical methods have led to more robust analyses of prosodic features and linguistic features in general. New approaches to the analysis of f0 tunes are constantly evolving and will change the future of prosodic analysis. Cumbersome manual annotations of tonal targets and error-prone automatic approaches will be replaced by more efficient and faster processes. Cangemi et al. (2019), for instance, propose a new approach which is not bound to any theoretical assumptions such as the AM approach that takes into consideration the shape of f0 patterns and text by means of periodic energy curves. In addition to advances in statistical methods and automatic labeling of data, advances in speech technology and language learning apps are to be expected. The development and design of speech learning software will be a challenge for the new decade. The prospect of providing learners with software that is able to give detailed feedback on pronunciation and prosody serves as a continuous incentive for future research.

[85] Jane Setter (2016) in her presentation on the paper by Mok et al. (2016) contends that fall-rises are avoided on purpose in order for them not to sound too British. The recording is available online.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 489-509.

Abrahamsson, N., & Hyltenstam, K. (2009). Age of L2 acquisition and native-likeness: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249-306.

Abu-Al-Makarem, A., & Petrosino, L. (2007). Reading and spontaneous speaking fundamental frequency of young Arabic men for Arabic and English languages: A comparative study. *Perceptual and Motor Skills*, *105*, 572-580.

Acton, W., Baker, A., Burri, M., & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In. J. Levis, & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 234-244). Aug. 2012. Ames, IA: Iowa State University.

Aguado-Jiménez, P., Pérez-Paredes, P., & Sánchez, P. (2012). Exploring the use of multidimensional analysis of learner language to promote register awareness. *System*, *40*(1), 90-103.

Aguilar, L., Downey, G., Krauss, R., Pardo, J., Lane, S., & Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: Both partners' rejection sensitivity matters. *Journal of Personality*, *84*(2), 165-177.

Algeo, J. (2010). *The origins and development of the English language*. Boston: Wadsworth (Cengage Learning).

Allan, S. (1990). The rise of New Zeland intonation. In A. Bell, & J. Holmes (Eds.), *New Zealand Ways of Speaking English* (pp. 115-128). Clevedon: Multilingual Matters.

Altenberg, B. (1987). *Prosodic patterns in spoken English*: *Studies in the correlation between prosody and grammar for text-to-speech conversation*. Lund: Lund University Press.

Anderson-Hsieh, J., Johnson R., & Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*; *42*(4), 529-555.

Andreeva, B., & Barry, W. J. (2012). Fine phonetic detail in prosody. Cross-language differences need not inhibit communication. In A. Steube, & O. Niebuhr (Eds.), *Understanding prosody: The role of context, function and communication* (pp. 259-288). De Gruyter.

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Oleskowicz-Popiel, M., & Trouvain, J. (2014a). Comparison of pitch range and pitch variation in Slavic and Germanic languages. In *Proceedings to the 7th Speech Prosody Conference, May 20-23 2014*, N. Campbell, D. Gibbon, & D. Hirst (Eds.), 776-780. Trinity College Dublin, Ireland. <https://www.isca-speech.org/archive/SpeechProsody_2014/abstracts/148.html>. (25 August 2017).

Andreeva, B., Demenko, G., Möbius, B., Zimmerer, F., Jügler, J., & Oleskowicz-Popiel, M. (2014b). Differences of pitch profiles in Germanic and Slavic languages. In *Proceedings of 15th Annual Conference of the International Communication Association (INTERSPEECH), September 14-18 2014*, 1307-1311. Singapore. <https://www.isca-speech.org/archive/interspeech_2014/i14_1307.html>. (25 August 2017).

Andrew, M. (2011). 'The real world': lived literacy practices and cultural learning from community placement. *Australian Journal of Language and Literacy*, *34*(2), 219-235.

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., & Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP-2002), Denver, Colorado, USA, September 16-20 2002,* J. Hansen, & B. Pellom (Eds.), 2037-2040.

Anufryk, V. (2008). Effects of Language Ability on Phonetic and Phonological Variation of English Intonation by German Speakers. In *8th International Seminar on Speech Production*, Strasbourg, France, December 8-12, 2008, R. Sock, S. Fuchs, & Y. Laprie (Eds.), 233-235. <https://issp2008.loria.fr/proceedings.html>. (02 September 2018).

Anufryk, V. (2012). *Intonational variation and pronunciation aptitude*. PhD dissertation. University Stuttgart.

Anufryk, V., & Dogil, G. (2009). On prosodic variation and pronunciation aptitude: A case study of German speakers of English. *The Journal of the Acoustical Society of America*, *125*(4), 2757. https://doi.org/10.1121/1.4784641

Anufryk, V., Jilka, M., & Dogil, G. (2008). Prosodic variation in L2: A case study of Germans speaking English. In *Proceedings of the ISCA Tutorial and Research Workshop on Experimental Linguistics (ExLing 2008)*, *August 25-27 2008,* 21-24. Athens, Greece. <https://www.isca-speech.org/archive/exling_2008/papers/exl8_021.pdf>. (23 August 2018).

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. In O.-S. Bohn, & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 282-297). John Benjamins.

Archibald, J. (1993a). The learnability of English metrical parameters by Spanish speakers. *International Review of Applied Linguistics*, *31*, 129-42.

Archibald, J. (1993b). Metrical phonology and the acquisition of L2 stress. In F. Eckman (Ed.), *Confluence: the interaction of linguistics, second language acquisition, and speech pathology* (pp. 37-48). Amsterdam: Benjamins.

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, *10*, 215-240.

Archibald, J. (1997). The acquisition of second language phrasal stress: A pilot study. In S. J. Hannahs, & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 263-281). Amsterdam: Benjamins.

Archibald, J. (2018). Advanced-level L2 phonology. In P. A. Malvoth, & A. G. Benati (Eds.), *The Handbook of Advanced Proficiency in Second Language Acquisition* (pp. 241-263). John Wiley and Sons.

Armstrong, M. E., Piccinini, P. E., & Ritchart, A. (2015). The phonetics and distribution of non-question rises in two varieties of American English. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), 0927. Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0927.pdf>. (04 September 2018).

Arvaniti, A. (2011). The representation of intonation. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rise (Eds.), *Companion to Phonology* (pp. 1-24). Wiley-Blackwell.

Arvaniti, A. (2017). The Autosegmental-Metrical model of intonational phonology. In S. Shattuck-Hufnagel, & J. Barnes (Eds.), *Prosodic Theory and Practice* (pp. 1-45). MIT Press.

Arvaniti, A., & Atkins, M. (2016). Uptalk in Southern British English. In *Proceedings of Speech Prosody, May 31-June 3 2016*, J. Barnes, & A. Brugos (Eds.), 153-157. Boston <https://www.isca-speech.org/archive/SpeechProsody_2016/pdfs/77.pdf>. (02 October 2018).

Arvaniti, A., & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 84-117). Oxford, UK: Oxford University Press.

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole, & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 547-576). Mouton de Gruyter: Berlin.

Arvaniti, A., & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, *26*, 43-74.

Asher, J. J., & Garcia, R. (1969). The optimal age to learn a foreign language. *Modern Language Journal*, *53*, 334-341.

Astruc, L., & del Mar Vanrell, M. (2016). Intonational phonology and politeness in L1 and L2 Spanish. *Probus*, *28*(1), 91-118.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of f0: Evidence from German. *Journal of Phonetics, 32,* 177-197.

Auran, C., & Hirst, D. (2004). Anaphora, connectives and resetting: prosodic and pragmatic parameters interactions in the marking of discourse structure. In *Proceedings of Speech Prosody*, *March 23-26, 2004, Nara, Japan,* B. Bel, & I. Marlien (Eds.), 259-262. <https://www.isca-speech.org/archive/sp2004/sp04_259.html>. (05 July 2018).

Ayers, G. M. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics, 44*, 1-49.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437-456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231-248.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123-150.

Backman, N. (1979). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin Utrecht*, *4*, 239-265.

Bacon, S. M. C. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in second-language listening. *Modern Language Journal*, *76*, 160-78.

Bacon, S. M. C., & Finnemann, M. D. (1992). Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input. *Language Learning*, *42*(4), 471-495.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. 2nd edition. Singular Publishing Group.

Barker, F. (2010). How can corpora be used in language testing? In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge Handbook of CorpusLinguistics* (pp. 633-645). Abingdon, UK: Routledge.

Barry, A. (2008). The *Form*, *Function And Distribution Of High Rising Intonation: a comparative study of HRT in Southern California and Southern British English.* Saarbrücken: Verlag Dr. Müller.

Barry, W. J. (2007). Rhythm as an L2 problem: How prosodic is it? In J. Trouvain, & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching* practice (pp. 97-120). Berlin: Mouton de Gruyter.

Barry, W., & Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. *Journal of the International Phonetic Association*, *31*(1), 51-66.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. doi:10.18637/jss.v067.i01.

Batliner, A., Buckow, J., Huber, R., Warnke, V., Nöth, E., & Niemann, H. (2001). Boiling down prosody for the classification of boundaries and accents in German and English. In *Eurospeech 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, September 3-7 2001,* 2781-2784. Aalborg, Denmark. <https://www.isca-speech.org/archive/eurospeech_2001/e01_2781.html>. (05 June 2018).

Baugh, Albert C., & Cable, T. (2013). *A History of the English Language*. 6th edition. London: Routledge.

Baumann, S., & Grice, M. (2006). The Intonation of Accessibility. *Journal of Pragmatics*, *38*, 1636-1657.

Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3,* 255-309.

Beckman, M. E., & Venditti, J. J. (2011). Intonation. In J. Goldsmith, J. Riggle & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 485-532). Blackwell Publishing.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9-54). Oxford University Press.

Beckman, M., & Ayers, G. (1997). Guidelines for ToBI labeling, version 3.0. Linguistics Department, Ohio State University.

Beckman, M., Díaz-Campos, M., McGory, J. T., & Morgan, T.A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, *14*, 9-36.

Bell, A., & Johnson, G. (1997). Towards a sociolinguistics of style. *University of Pennsylvania Working Papers in Linguistics*, *4*, 1-21.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation*, *6*(4), 802-813.

Beňuš, Š., Reichel, U. D., & Mády, K. (2014). Modeling accentual phrase intonation in Slovak and Hungarian. In L.Veselovská, & M. Janebová (Eds.), *Complex Visibles Out There. Olomouc modern language series 4* (pp. 677-689). Olomouc, Czech Republic: Palacký University.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204). Timonium: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro, & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Bever, T. G. (1981). Normal acquisition processes explain the critical period for language learning. In K. Diller (Ed.), *Individual differences and universal in language learning aptitude* (pp. 176-198). Rowley, MA: Newbury House.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2002). *Longman Grammar of Spoken and Written English*. Harlow, England; New York: Pearson Education ESL.

Bielska, J. (2009). Gender differences in strategy use and classroom task preferences of Polish EFL learners. In M. Navarro Coy (Ed.), *Practical Approaches to Foreign Language Teaching and Learning* (pp. 113-127). Bern: Peter Lang.

Biersack, S. (2002). Systematische Aussprachefehler deutscher Muttersprachler im Englischen: eine phonetisch-phonologische Bestandaufnahme (Systematic pronunciation mistakes of German natives speaking English). MA thesis published in *Forschungsberichte des Instituts für Phonetik und SprachlicheKommunikation der Universität München (FIPKM), 39*, 37-130.

Biersack, S., Kempe, V., & Knapton, L., (2005). Fine-tuning speech registers:a comparison of the prosodic features of child-directed and foreigner directed speech. In *Proceedings of 9th European Conference on Speech Communication and Technology Eurospeech, September 4-8 2005*, 2401-2404. Lisbon, Portugal. <https://www.isca-speech.org/archive/interspeech_2005/i05_2401.html>. (25 August 2017).

Bigi, B. (2015). SPPAS - Multilingual Approaches to the Automatic Annotation of Speech. *The Phonetician*, *International Society of Phonetic Sciences*, *111-112*, 55-69.

Bilious, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication*, *8*(3-4), 183-194.

Birdsong, D. (1999). Introduction: Whys and why nots of the Critical Period Hypothesis for Second Langauge Acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 1-22). Mahwah, NJ: Erlbaum.

Birdsong, D. (2005). Native-likeness and nonnative-likeness in L2A research. *International Review of Applied Linguistics*, *43*, 319-328.

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, *35*(4), 374-392.

Bissiri, M. P. (2013). Glottalization in German-accented English in relationship to phrase boundaries. In D. Methnert, U. Kordon, & M. Wolff (Eds.), *Systemtheorie Signalverarbeitung Sprachtechnologie*. *Rüdiger Hoffman zum 65. Geburtstag. Studientexte zur Sprachkommunikation Band* 68 (pp.234-240). Dresden: TUD press.

Bissiri, M. P., & Volín, J. (2010). Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English. In *20th Czech-German Workshop on Speech Processing*, *September 20 2010*, R. Vích (Ed.), 23-28. Institute of Photonics and Electronics, Academy of Sciences of the Czech Republic, Prague.

Bissiri, M. P., Lecumberri, M. L., Cooke, M., & Volín, J. (2011). The role of word-initial glottal stops in recognizing English words. In *Proceedings of Interspeech*, *12th Annual Conference of the International Speech Communication Association, August 27-31 2011*, P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini, 165-168. Firenze, Italy. <https://www.isca-speech.org/archive/interspeech_2011/i11_0165.html>. (02 July 2019).

Bitouk, D., Ragini, V., & Nenkova, A. (2010). Class-level spectral features for emotion recognition. *Speech Communication*, *52*(7-8), 613-625.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14*, 359-375.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford, & M. Sharwood Smith (Eds.), *Grammar and second language teaching* (pp. 19-30). Rowley, MA: Newbury House.

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass, & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 41-68). New York: Cambridge University Press.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3-49.

Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.03, retrieved 1 September 2019 from http://www.praat.org/

Bolinger, D. (1946). The intonation of quoted questions. *Quarterly Journal of Speech, 32*(2), 197-202.

Bolinger, D. (1958). A theory of pitch accent in English. *Word*, *14*, 109-149.

Bolinger, D. (1972). Around the edge of language: Intonation. In D. Bolinger (Ed.), Intonation (pp.19-29). Harmondsworth: Penguin.

Bolinger, D. (1986). *Intonation and Its Parts: melody in spoken English.* London: Arnold.

Bolinger, D. (1989). *Intonation and its Uses.* London: Edward Arnold

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 133-159). Mahwah, NJ: Erlbaum.

Bongaerts, T., Mennen, S., & Van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica, 54*(2)*,* 298-308.

Bongaerts, T., Planken B., & Schils, E. (1995). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton, & Z. Lengyel (Eds.), T*he age factor in second language acquisition* (pp. 30-50). Clevedon, U.K.: Multilingual Matters.

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*, 447-465.

Bourhis R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119-135). New York: European Association of Experimental Social Psychology.

Bowden, H., Sanz, C., & Stafford, C. (2005). Individual differences: Age, sex, working memory, and prior knowledge. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 105-140). Georgetown University Press.

Bowen, D. (1956). A comparison of the intonation patterns of English and Spanish. *Hispania*, *39*(1), 30-35.

Boyle, J. P (1987). Sex differences in listening vocabulary. *Language Learning*, *37*(2). 273-284.

Brabcová, K., & Skarnitz , R. (2018). Foreign or native-like? The attitudes of Czech EFL learners towards accents of English and their use as pronunciation models. *Studie z aplikované lingvistiky - Studies in Applied Linguistics*, 9(1), 38-50. Charles University in Prague: Faculty of Arts. Czech Republic.

Bradford, B. (1997). Upspeak in British English. *English Today*, *13*(3), 29-36.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729.

Brand, C., & Kämmerer, S. (2006). The Louvain International Database of Spoken English Interlanguage (LINDSEI): Compiling the German component. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods* (pp. 127-140). Frankfurt: Peter Lang.

Braun, A. (1995). Fundamental frequency - How speaker-specifc is it? *Studies in Forensic Phonetics, BEIPHOL 64,* 9-23.

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Brazil, D., Coulthard, M. & Johns, C. (1980). Discourse intonation and language teaching. Harlow: Longman.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, *34*, 383-398.

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language and Communication*, *22*, 171-18.

Britain, D. (1998). Linguistic change in intonation: The use of high-rising terminals in New Zealand English. In P. Trudgill, & J. Cheshire (Eds.), *The sociolinguistics reader* 1: *Variation and multilingualism* (pp. 213-239). London: Arnold.

Britain, D., & Newman J. (1992). High rising terminals in New Zealand English. *Journal of the International Phonetic Association*, *22*(1-2), 1-11.

Britain, D., (1992). Linguistic change in intonation: the use of high rising terminals in New Zealand English. *Language Variation and Change*, *4*, 77-104.

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler, & D. R. Ladd (Eds.), *Prosody: Models and* Measurements (pp. 67-77). Berlin: Springer.

Brown, P., & Levinson, S. C. (1987). *Politeness. Some universals in language usage. Studies in Interactional Sociolingustics 4*. Cambridge: Cambridge University Press.

Brugos, A., Breen, M., Veilleux, N., Barnes, J., & Shattuck-Hufnagel, S. (2018). Cue-based annotation and analysis of prosodic boundary events. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 245-249. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/246.html>. (17 July 2018).

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*, 257-277.

Bryant, P. (1980). Australian questioning intonation: An addition to speakers' response-seeking repertoire. BA (honours) thesis, Australian National University.

Buck, G. (2016). Systematic transfer of 'uptalk' from English (L1) to Spanish (L2) for native English speakers: A developmental study. Bachelor's thesis. University of Michigan.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with L2 vowel perception performance in adult L2 learners. *Studies in Second Language Acquisition*, *33*, 433-461.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*(3), 643-664.

Burdin, R. St. & Tyler, J. (2018). Rises inform, and plateaus remind: Exploring the epistemic meanings of "list intonation" in American English. *Journal of Pragmatics*, *136*, 97-114.

Busà, M. G., & Urbani, M. (2011). A cross linguistic analysis of pitch range in English L1 and L2. In *ICPhS XVII, August 17-21 2011,* 380-383. Hong Kong. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Busa/Busa.pdf>. (03 August 2017).

Buysse, L. (2017). The pragmatic marker *you know* in learner Englishes. *Journal of* Pragmatics, 121, 40-57.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does first language maintenance hamper native-likeness in a second language? *Studies in Second Language Acquisition*, *34*(2), 215-241.

Cacioppo, J. T., & Petty, R. E. (1980). Sex differences in influenceability: Towards specifying the underlying processes. *Personality and Social Psychology Bulletin*, *6*(4), 651-656.

Callies, M. (2008). Easy to understand but difcult to use? Raising constructions and information packaging in the advanced learner variety. In G. Gilquin, M. Belén Diez Bedmar, & S. Papp (Eds.), *Linking Contrastive and Learner Corpus Research* (pp. 201-226). Amsterdam: Rodopi.

Callies, M. (2009). *Information Highlighting in Advanced Learner English: The syntax-pragmatics interface in second language acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing.

Callies, M., & Szczesniak, K. (2008). Argument realization, information status and syntactic weight - A learner-corpus study of the dative alternation. In P. Grommes, & M. Walter (Eds.), *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung* (Linguistische Arbeiten, Band 520) (pp.165-187). Tübingen: Niemeyer.

Cameron, D. (2011). Sociophonetics and sexuality: Discussion. *American Speech*, *86*(1), 98-103.

Cangemi, F., Albert, A., & Grice, M. (2019). Modelling intonation: Beyond segments and tonal targets. In *Proceedings of the 19th International Congress of Phonetic Sciences* (*ICPhS)*, *August 5-9 2019*, S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), 572-576. Melbourne, Australia. <http://intro2psycholing.net/ICPhS/papers/ICPhS_621.pdf>. (14 August 2019).

Carli, L. L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology*, *59*(5), 941-51.

Carlsen, C. (2012). Proficiency level - a fuzzy variable in computer learner corpora. *Applied Linguistics*, *33*(2), 161-183.

Carrie, E. (2017). British is professional, American is urban: Attitudes towards English reference accents in Spain. *International Journal of Applied Linguistics*, *27*(2), 427-447.

Carver, R. H., & Gradwohl Nash, J. (2011). Doing Data Analysis with SPSS. Version 18.0. Duxbury.

Cauvin, É. (2013). Intonational phrasing as a potential indicator for establishing prosodic learner profiles. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use - Proceedings* 1 (pp. 75-88). Louvain-la-Neuve: Presses universitaires de Louvain.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, *October 3-6 1996*, 2175-2179. Philadelphia, PA, USA. <https://www.isca-speech.org/archive/icslp_1996/i96_2175.html>. (09 July 2018).

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching Pronunciation: a course book and reference guide*. Cambridge: Cambridge University Press.

Cenoz, J., & García Lecumberri, M. L. (1999). The acquisition of English pronunciation: learners' views. *International Journal of Applied Linguistics*, *9*(1), 3-15.

Chafe, W. L. (1979). The flow of thought and the flow of language. In T. Givon (Ed.), *Discourse and syntax. Syntax and semantics* (pp. 159-181), 12. New York: Academic Press.

Chafe, W. L. (1980). *The Pear Stories. Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.

Chafe, W. L. (1988). Linking intonation units in spoken English. In J. Haiman & S. A. Thompson (Eds.), *Clause Combining in Grammar and Discourse* (pp. 1-27). Amsterdam, Philadelphia: Benjamins.

Chafe, W. L. (1994). *Discourse, consciousness and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: Chicago University Press.

Chamonikolasová, J. (2017). *Intonation in English and Czech Dialogues*. Brno: Masarykova univerzita.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910.

Chen, A. (2003). Language Dependence in Continuation Intonation. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, *3-9 August 2003,* M. Solé,

D. Recasens and J. Romero (Eds.), 1069-1072. Barcelona. <https://www.internationalphoneticassociation. org/icphs-proceedings/ICPhS2003/ papers/p15_1069.pdf>. (06 August 2018).

Chen, A. (2009). Perception of paralinguistic intonational meaning in a second language. *Language Learning*, *59*(2), 367-409.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, *120*, 1926-1939.

Ching, M. K. L. (1982). The question intonation in assertions. *American Speech*, *57*, 95-107.

Chipere, N.(2013). Sex differences in phonological awareness and reading ability. *Language Awareness*, *23*(3), 1-15.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning and Technology*, *2*(1), 74-93.

Chun, D. M. (2002). *Discourse Intonation in L2: from theory and research to practice*. Amsterdam: John Benjamins.

Chun, D. M., Hardison, D. M., & Pennington, M. C. (2008). Technologies for prosody in context. Past and future of L2 research and practice. In J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*, *Studies in Bilingualism 36* (pp. 323-346). John Benjamins.

Clark, H. H. (1994). Discourse in production. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 985-1021). San Diego, CA: Academic Press..

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*, 73-111.

Clopper, C. G., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication, 48*, 633-644.

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal Phonetics*, *39*(2), 237-245.

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1-15.

Coates, J. (2004). *Women, Men, and Language: A Sociolinguistic Account of Gender Differences in Language*. Pearson Education.

Cogo, A., & Dewey, M. (2006). Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. In A. Mauranen, & M. Metsa-Ketela (Eds.), *Nordic Journal of English Studies: Special Issue on ELF*, 59-93.

Cohen, A., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua*, *19*, 177-192.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Meassurement*, *20*(1), 37-46.

Colantoni, L., Cuza, A., & Mazzaro, N. (2016a). Task-related effects in the prosody of Spanish heritage speakers and long-term immigrants. In M. E. Armstrong, N. Henrikesen, & M. del Mar Vanrell (Eds.), *Intonational Grammar in Ibero-Romance: Approaches across linguistic subfields. Issues in Hispanic and Lusophone Linguistics* 6 (pp. 1-24). John Benjamins.

Colantoni, L., Klassen, G., Patience, M., Radu, M., & Tararova, O. (2015). Increasing context: L2 production of English intonation by L1 Spanish and L1 Mandarin speakers.

In *Proceedings of the 18th Congress of the International Phonetic Association (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), 0813. Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0813.pdf>. (07 July 2018).

Colantoni, L., Klassen, G., Patience, M., Radu, M., & Tararova. (2016b). Task-effects in the L2 perception and production of English sentence-types by L1 Spanish speakers. *Proceedings of Speech Prosody 8. Boston: Boston University.*

Conley, J. M., O'Barr, W. M., & Lind, A. E. (1978). The power of language: Presentational style in the courtroom. *Duke Law Journal*, *6*,1375-1399.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, *33*(2), 185-209.

Cook, V. (ed.) (2002). *Portraits of the L2 user*. Clevedon, UK: Multilingual Matters.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics in Language Teaching*, *5*(4), 161-170.

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes*, *23*, 528-556.

Costrich, N., Feinstein, J., & Kidder, L. (1975). When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology*, *11*, 520-530.

Council of Europe (2018). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors. Strasbourg.

Couper-Kuhlen, E., & Selting, M. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen, & M. Selting (Eds.), *Prosody in Conversation: Interactional* Studies (pp. 11-56). Cambridge: Cambridge University Press.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Cruttenden, A. (1994). Rises in English. In J. Lewis (Ed.), *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor* (pp. 155-73). London: Routledge.

Cruttenden, A. (1997). *Intonation*. Cambridge: Cambridge University Press. 2nd edition.

Cruz-Ferreira, M. (1983). Non-native comprehension of intonation patterns in Portuguese and in English. Unpublished Ph.D. thesis, University of Manchester.

Crystal, D. (1969). *Prosodic systems and intonation in English.* Cambridge: Cambridge University Press.

Crystal, D. (1986). Prosodic development. In P. J. Fletcher, & M. Garman (Eds.), *Studies in First Language Development* (pp. 174-197). New York: Cambridge University Press.

Crystal, D. (1995). *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.

Crystal, D. (2003a). *English as a Global Language*. Cambridge University Press.

Crystal, D. (2003b). *A dictionary of linguistics and phonetics.*6th edition.Oxford: Blackwell Publishing.

Cucchiarini, C., Strik H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, *107*, 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 111, 2862-2873.

Culpeper, J., Mackey, A., & Taguchi, N. (2018). *Second language pragmatics. From theory to research*. New York: Routledge.

Cutler, A., & Swinney, D. A. (1987). Prosody and the development of comprehension. *Journal of Child Language*, *14*(1), 145-167.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2): 141-201.

Czech Statistical Office (2019). *Data - počet cizinců*. Available from: <https://www.czso.cz/documents/11292/27914491/1612_c01t04.pdf>. (22 September 2019).

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition*, *7*(1), 73-79.

Dalton-Puffer, C., Kaltenboeck, G., & Smit, U. (1997). Learner attitudes and L2 pronunciation in Austria. *World Englishes*, *16*(1), 115-28.

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, *5*(1), 85-96.

Daneš, F. (1957). *Intonace a věta ve spisovné češtině* [Intonation and the sentence in standard Czech]. Praha: Academia.

Dankovičová, J. (1998). *The Linguistic Basis of Articulation Rate Varation in Czech*. PhD dissertation. Oxford University.

Dascal, M. (1999). Introduction: Some questions about misunderstanding. *Journal of Pragmatics*, *31*, 753-762.

Davies, A. (2003). *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters.

De Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*(4), 331-350.

De Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, *16*(1), 71-77.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL)*, New Series *2*, 225-246.

De Leeuw, E. (2019). Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades. *Journal of Phonetics*, *73*, 24-39.

De Leeuw, E., & Celata, C. (2019). Plasticity of native phonetic and phonolgical domains in the context of bilingualism. *Journal of Phonetics*, *73*, 88-93.

De Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*, *16*, 101-116.

De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014). Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, *58*,11-34.

DeKeyser, R. M., Alf-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*(3), 413-438.

Delattre, P. (1963). Comparing the prosodic features of English, German, Spanish and French. *International Review of Applied Linguistics*, *1*, 193-210.

Delattre, P., Poenack, E., & Olsen, C. (1965). Some characteristics of German Intonation for the Expression of Continuation and Finality. *Phonetica*, *13*, 134-161.

Derwing, T. M, & Munro, M. J. (1997). Accent, intelligibility and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1-16.

Derwing, T. M., & Munro, Murray J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163-185.

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, *13*(1), 1-17.

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2013). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*(1), 22-44.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393-410.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533-557.

Dewaele, J.-M., MacIntyre, P., Boudreau, C., & Dewaele, L. (2016). Do girls have all the fun? Anxiety and enjoyment in the Foreign Language Classroom. *Theory and Practice of Second Language Acquisition*, *2*(1), 41-63.

Di Gioacchino, M., & Crook Jessop, L. (2011). Uptalk - Towards a quantitative analysis. *Toronto Working Papers in Linguistics, 33*, 1-16.

Diehl, R. L., Lindblom B., Hoemeke K. A., & Fahey, R. P. (1996). On explaining certain male-female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24*(2), 187-208.

Dietrich, R., Klein, W., & Noyau, C. (1995). *The Acquisition of Temporality in a Second Language*. Amsterdam: John Benjamins.

Dijkstra, C., Krahmer, E., & Swerts, M. (2006). Manipulating uncertainty - The contribution of different audiovisual prosodic cues to the perception of confidence. In *Proceedings of the International Conference of Speech Prosody*, *May 2-5 2006*, R. Hoffmann, & H. Mixdorff (Eds.). Dresden, Germany. <https://www.isca-speech.org/archive/sp2006/sp06_025.html>. (25 September 2019).

Dimos, K., Dick, L., & Dellwo, V. (2015). Perception of levels of emotion in speech prosody. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, *August 10-14 2015.* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0756.pdf>. (08 June 2018).

Dombrowski, E., & Niebuhr, O. (2005). Acoustic patterns and communicative functions of phrase-final rises in German: Activating and restricting contours. *Phonetica*, *62*, 176-195.

Dombrowski, E., & Niebuhr, O. (2010). Shaping phrase-final rising intonation in German. In *Proceedings of the 5th International Conference of Speech Prosody*, *May 11-14 2010*, 1-4. Chicago, IL, USA. <https://www.isca-speech.org/archive/sp2010/papers/sp10_788.pdf>. (20 August 2018).

Dörnyei, Z., & Clément, R. (2001). Motivational characteristics of learning different target languages: Results of a nationwide survey. In Z. Dornyei, & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 399-432). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Dragojevic, M., & Giles, H. (2016). I don’t like you because you’re hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, 42, 396-420.

Draxler, C. (2008). *Korpusbasierte Sprachverarbeitung: Eine Einführung*. Narr Studienbücher: Tübingen.

Du Bois, J. W., Schuetze-Coburn, S., Paolino, D., & Cummings, S. (1992). Discourse transcription, Santa Barbara Papers in Linguistics 4, Dept. of Linguistics, University of California, Santa Barbara.

Duběda, T. (2011). Towards an inventory of pitch accents for read Czech. *Slovo a slovesnost*, *72*(1), 3-12.

Duběda, T. (2014). Czech intonation: A tonal approach. *Slovo a slovesnost*, *2*, 83-98

Duběda, T., & Mády, K. (2010). Nucleus position within the intonation phrase: a typological study of English, Czech and Hungarian. In In *Proceedings of 11th Annual Conference of the International Speech Communication Association (INTERSPEECH), September 26-30 2010*, 126-129. Makuhari, Chiba, Japan. <https://www.phonetik.uni-muenchen.de/~mady/pub/dubeda_mady10.pdf>. (9 August 2019).

Dumont, A. (2017). The contribution of learner corpora to the substantiation of fluency levels. In P. De Haan, R. De Vries, & S. Van Vuuren (Eds.), *Language, Learners and Levels: Progression and* Variation (pp. 281-308). Louvain-la-neuve: Presses Universitaires de Louvain.

Dziubalska-Kołaczyk, K., & Przedlacka, J. (Eds.). (2008). English pronunciation models: a changing scene. 2nd edition. Hamburg: Peter Lang.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, *85*(1), 86-116.

Eckert, P., & Podesva, R. J. (2011). Sociophonetics and sexuality: Toward a symbiosis of sociolinguistics and laboratory phonology. *American Speech*, 86 (1), 6-13.

Eckman, F. R. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning*, 27(2), 315-330.

Eckman, F. R. (1996). A functional-typological approach to second language acquisition theory. In W. C. Ritchie, & T. K. Bhatia (Eds), *Handbook of second language acquisition*. New York: Academic Press.

Edelsky, C. (1979). Question intonation and sex roles. *Language in Society*, *8*, 15-32.

Edlund, J., Heldner, M., & Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proceedings of 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *September 6-10 2009*, 2779-2782. Brighton, United Kingdom. <https://www.isca-speech.org/archive/interspeech_2009/i09_2779.html>. (02 August 2019).

Education First (2011-2015). EF EPI: EF English Proficiency Index (five editions). Retrieved 1 September 2019 from http://www.ef.co.uk/epi/downloads/ .

Edwards, A, & Fuchs, R. (2018). A cluster analysis of attitudes to English in Germany and the Netherlands. *World Englishes*, *37*(4), 653-667.

Ehrman, M., & Oxford, R. (1989). Effects of sex differences, career choice, and psychological type on adult language learning strategies. *Modern Language Journal*, *72*(3), 253-265.

Elliott, R. E. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. T*he Modern* L*anguage Journal*, *79*, 356-371.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. (2008). *The Study of Second Language Acquisition.* 2nd edition. Oxford: Oxford University Press.

Elvira-García, W. (2016). La entonación en el aula de ELE. In N. Baulenas, & B. Montmany (Eds.), *Dossier del XXV encuentro práctico de profesores de ELE* (pp. 41-47). Barcelona: Difusión Publishing Company and International House.

Eriksson, A., & Heldner, M. (2015). The acoustics of word stress in English as a function of stress level and speaking style. In *Proceedings of 16th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *September 6-10 2015*, 41-45. Dresden, Germany. <https://www.isca-speech.org/archive/interspeech_2015/i15_0041.html>. (15 September 2018).

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, *2*(2), 62-76

Esser, J. (1978). Contrastive intonation of German and English. *Phonetica*, *35*, 41-55.

Estebas-Vilaplana, E. & Prieto, P. (2010). Castilian Spanish Intonation. In P. Prieto, & P. Roseano (Eds.), *Transcription of Intonation of the Spanish Language* (pp. 17-48). Munich, Lincom Europa.

Estebas-Vilaplana, E. (2008). Modelling final declarative intonation in English and Spanish. In T. Gibert, & L. Alba (Eds.), *Estudios de Filología Inglesa. Homenaje a la Dra. Asunción Alba Pelayo* (pp. 109-127). Madrid: UNED.

Estebas-Vilaplana, E. (2014). The evaluation of intonation: pitch range differences in English and Spanish. In G. Thompson, & L. Alba-Juez (Eds.), *Evaluation in Context - Pragmatics & Beyond New Series* (pp. 179-194). Amsterdam: John Benjamins.

Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica del español: una revisión del Sp_ToBI. *Estudios de Fonética Experimental*, *17*, 265-283.

Euler, S. (2014). Approaches to pronunciation teaching: History and recent development. In Jolanta Szpyra-Kozłowska, Ewa Guz, Piotr Steinbrich, Radosław Święciński (Eds.). Recent Developments in Applied Phonetics. 35-78.

Face, T. L. (2001). Focus and early peak alignment in Spanish intonation. *Probus: International Journal of Latin and Romance Linguistics*, *13*(2), 223-246.

Face, T. L. (2005). F0 peak height and the perception of sentence type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *2*(6), 49-65.

Feagin, C. (1997). The African contribution to southern states English. In C. Bernstein, T. Nunally, & R. Sabino, R. (Eds.), *Language Variety in the South Revisited*. (pp. 123-139). University of Alabama Press; Tuscaloosa, AL.

Ferguson, C. A. (1971). Absence of copula and the notion of simplicity: A study of normal speech, baby talk, foreigner talk and pidgins. In H. Dell Hymes (Ed.), *Pidginization and Creolization of Languages* (pp. 141-150). New York: Cambridge University Press.

Fernández, E. M. (2010). Reading aloud in two languages. The interplay of syntax and prosody. In Bill VanPatten & Jill Jegerski. *Research in Second Langauge Processing and Parsing. Language Acquisition and Language Disorders*, *53*, 297-320.

Fernández, J., & Staples, S. (2021). Pragmatic approaches. In N. Tracy-Ventura, & M. Paquot. *Routledge handbook of SLA and corpora*. Routledge.

Flaig, S., & Zerbian, S. (2016). Rising intonation in statements of German speakers of English as a foreign language. *The 7th conference on Tone and Intonation in Europe (TIE), September 1-3 2016,* Canterbury. <https://blogs.kent.ac.uk/tie-conference/files/2016/08/Flaig_Zerbian.pdf>. (05 June 2018).

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*, 70-79.

Flege, J. E. (1995). Second language speech learning theory, findings and problems. W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in cross-linguistic research* (pp. 233-277). Timonium: York Press.

Flege, J. E. (2009). Give input a chance! In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175-190). Bristol: Multilingual Matters.

Flege, J. E. (2016). The role of phonetic category formation in second language speech acquisition. *Eighth International Conference on Second Language Speech, June 10-12 2016*. Aarhus University, Denmark. < http://www.jimflege.com/files/Flege_Aarhus_talk_20162.pdf>. (02 October 2017).

Flege, J. E. (2018). How the revised Speech Learning Model (SLM-r) works for stop consonants. Conference Paper. Research Frontiers of Second Language Speech, April 14-15 2018, Nanjing University of Science and Technology, China. <http://jimflege.com/files/SLM_talk_Nanjing.pdf>. (08 August 2018).

Flege, J. E. (2019). A non-critical period for second-langauge learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A sound approach to language matters. In Honor of O.- S. Bohn* (pp. 501-541). Department of English, Aarhus University. Open Access e-book at Aarhus University Library.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived accent. *Journal of the Acoustical Society of America*, *91*, 370-389.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in second language acquisition*, *26*(1), 1-34.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997a). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437-470.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997b). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169-186.

Flege, J. E., MacKay, I- R.A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics, 23,* 567-598.

Flege, J. E., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134.

Fletcher, J. (2005). Compound rises and “uptalk” in spoken English dialogues. In *Proceedings of 9th European Conference on Speech Communication and Technology Eurospeech*, *September 4-8 2005*, 1381-1384. Lisbon, Portugal. <https://www.isca-speech.org/archive/interspeech_2005/i05_1381.html>. (04 April 2018).

Fletcher, J., & Loakes, D. (2006). Patterns of rising and falling in Australian English, In *Proceedings of the 11th Australian Conference on Speech Science and Technology*, *December 6-8 2006*, P. Warren, & C. Watson (Eds.), 42-47. University of Auckland, New Zealand.

Fletcher, J., Grabe, E.,& Warren, P. (2005). Intonational variation in four dialects of English: the high rising tone. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 390-409). Oxford: Oxford University Press.

Flynn, S., & Manuel, S. (1991). Age-dependent effects in language acquisition: An evaluation of “Critical Period” hypotheses. In L. Eubank (Ed.), *Point counterpoint. Universal Grammar in the second language* (pp. 117-145). John Benjamins.

Ford, C. E. (1993). *Grammar and Interaction: Adverbial Clauses in American English Conversations*. Cambridge: Cambridge University Press.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 134-184). Cambridge: Cambridge University Press.

Foster, P., & Snyder Ohta, A. (2005). Negotiation for Meaning and Peer Assistance in Second Language Classrooms. *Applied Linguistics*, *26*(3). 402-430.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, *59*(4), 866-896.

Fougeron, C., & Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *Journal of Phonetics*, *26*, 45-69.

Fought, C. (2003). *Chicano English in Context*. Basingstoke: Palgrave Macmillan.

Fox, A. (1981). Fall-rise intonations in German and English. In C. V. J. Russ (Ed.), *Contrastive aspects of English and German* (pp. 55-72). Heidelberg: Julius Groos.

Fox, A. (1984). *German Intonation: An Outline*. Oxford: Clarendon Press.

Fox, R.A., Jacewicz, E., & Hart, J. (2013). Pitch pattern variations in three regional varieties of American English. In *Proceedings of 14th Annual Conference of the International Communication Association (INTERSPEECH), August 25-29 2013*, 123-127. Lyon, France. <https://www.isca-speech.org/archive/interspeech_2013/i13_0123.html>. (05 Januar 2018).

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*, 275-301.

Friedberg, H., Litman, D., & Paletz, S. (2012). Lexical entrainment and success in student engineering groups. *The 4th* IEEE Workshop on *Spoken Language Technology (SLT), December 2-5 2012,* 404-409. Miami, Florida, USA.

Friginal, E., Lee, J., Polat, B., & Roberson, A. (2017). *Exploring Spoken English Learner Language from Corpora: Learner Talk*. London: Palgrave McMillan.

Fuchs, R. (2018). Pitch range, dynamism and level in postcolonial varieties of English: A comparison of educated Indian English and British English. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 893-897. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/167.html>. (17 June 2018).

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 39-55). New York: Springer-Verlag.

García Lecumberri, M. L., Cooke, M., & Wester, M. (2017). A bi-directional task-based corpus of learners‘ conversational speech. *International Journal of Learner Corpus Research*, *4*(2), 175-195.

García, O. (2009). Emergent Bilinguals and TESOL: What's in a Name? *TESOL Quarterly*, *43*, 322-326.

Gass, S. M. (2013). *Second Language Acquisition: An Introductory Course*. Routledge.

Gass, S. M., & Varonis, E. (1985). Variation in native speaker speech modification to nonnative speakers. *Studies in Second Language Acquisition, 7,* 37-58.

Gilbert, J. B. (2008). *Teaching Pronunciation: Using the Prosody Pyramid*. Cambridge: Cambridge University Press.

Giles, H., & Powesland, P. (1997). Accommodation Theory. In N. Coupland; & A. Jaworsky (Eds.), *Sociolinguistics: A Reader and Coursebook* (pp. 232-239). Palgrave: New York.

Giles, H., Coupland, J., & Coupland, N. (1991). Accommodation Theory: Communication, Context, and Consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), Contexts of Accommodation (pp. 1-68). New York, NY: Cambridge University Press.

Gilles, P. (2005). Regionale Prosodie im Deutschen. Varibilität in der Intonation von Abschluss und Weiterweisung. In P. Giles (Ed.), *Linguistik - Impulse & Tendenzen 6*. De Gruyter.

Gilquin, G., De Cock, S., & Granger, S. (Eds.). (2010). LINDSEI: Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM. Louvain-la-Neuve: Presses universitaires de Louvain.

Goldman, J. - P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. In *Proceedings of 12th Annual Conference of the International Speech Communication Association (INTERSPEECH), August 27-31 2011*, P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini, 3233-3236. Firenze, Italy. <https://www.isca-speech.org/archive/interspeech_2011/i11_3233.html>. (02 May 2018).

González Ardeo, J. M. (2003) Attitude towards English and ESP acquisition as an L2 or L3 at university. *Ibérica*, *6*, 109-33.

Götz, S. (2013). *Fluency in Native and Nonnative English Speech*. Amsterdam: Benjamins.

Götz, S. (2019a). Do learning context variables have an effect on learners‘ (dis)fluency? Language-specific vs. universal patterns in advanced learners’ use of filled pauses. In L. Degand, G. Gilquin, L. Meurant, & A.C. Simon (Eds.), *Fluency and Disfluency across Languages and Language Varieties* (pp. 177-196). Louvain-la-neuve: Presses Universitaires de Louvain.

Götz, S. (2019b). Filled pauses across proficiency levels, L1s and learning context variables - A multivariate exploration of the *Trinity Lancaster Corpus Sample*. *International Journal of Learner Corpus Research* 5:2. 159-180.

Götz, S., & Mukherjee, J. (2018). Investigating the effect of the study abroad variable on learner output: A pseudo-longitudinal study on spoken German learner English. In V. Brezina, & L- Flowerdew (Eds.), *Learner Corpus Research: New perspectives and* applications (pp. 47-65). London and New York: Bloomsbury.

Grabe, E. (1997). Comparative Intonational Phonology: English and German. In *Intonation: Theory, Models, and Applications*, *September 18-20 1997*, A. Botinis (Ed.), 157-160.Athens, Greece. <https://www.isca-speech.org/archive_open/int_97/inta_157 .html>. (05 April 2017).

Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. PhD, Max-Planck-Institut for Psycholinguistics and University of Nijmegen.

Grabe, E. (2002). Variation adds to prosodic typology. In *Proceedings of Speech Prosody*, *April 11-13 2002,* B. Bel, & I. Marlin (Eds.), 127-132. Aix-en-Provence, France: Laboratoire Parole et Langage. <https://www.isca-speech.org/archive/sp2002/sp02_127.html>. (15 May 2017).

Grabe, E. (2004). Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles, & J. Peters (Eds.), *Regional Variation in Intonation* (pp. 9-31). Linguistische Arbeiten. Tuebingen, Niemeyer.

Grabe, E., & Post, B. (2002). Intonational Variation in English. In *Proceedings of Speech Prosody*, *April 11-13 2002,* B. Bel, & I. Marlin (Eds.), 343-346. Aix-en-Provence, France: Laboratoire Parole et Langage. <https://www.isca-speech.org/archive/sp2002/sp02_343.html>. (15 May 2017).

Grabe, E., Kochansky, G., & Coleman, J. (2005). The intonation of native accent varieties in the British Isles: Potential for miscommunication? In K. Dziubalska-Kolaczyk, & J. Przedlacka (Eds.), *English pronunciation models: a changing scene. Linguistics Insights: Studies in Language and Communication* (pp. 311-338). Bern: Peter Lang.

Grabe, E., Post, B., & Nolan, F. (2001). The IViE Corpus. Cambridge: University of Cambridge: Department of Linguistics. Available at: http://www.phon.ox.ac.uk/files/apps/IViE/index.php

Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*, 161-185.

Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, *46*(4), 375-401.

Grabe, E.; Nolan, F., & Farrar, K. (1998). IViE - A comparative transcription system for intonational variation in English. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP), November 30 - December 4 1998,* paper 0099. Sydney, Australia. <https://www.isca-speech.org/archive/icslp_1998/i98_0099.html>. (06 August 2017).

Gráf, T. (2015). *Accuracy and fluency in the speech of the advanced learner of English*. Phd.

Gráf, T. (2017). *Corpus LINDSEI_CZ*. Praha: FF UK.

Gráf, T. (2019). Speech rate revisited: The effect of task design on speech rate. In S. Götz, & J. Mukherjee (Eds.), *Learner Corpora and Language Teaching* (pp. 175-189). Amsterdam: John Benjamins.

Gráf, T., & Huang, L.-F. (2019). Repeats in native and learner English. In L. Degand, G. Gilquin, L. Meurant, & A. Catherine Simon (Eds.), *Fluency and Disfluency across Languages and Language Varieties. Corpora and Language in Use - Proceedings 4* (pp. 219-241). Louvain-la-Neuve: Presses universitaires de Louvain.

Graham, C., & Post, B. (2018). Second language acquisition of intonation: Peak aligment in American English. *Journal of Phonetics*, *66*, 1-14.

Granena, G., & Long, M. H. (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311-343.

Granger, S. (1996). From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In K. Aijmer (Ed.), *Languages in Contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund University Press: Lund.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-18). London: Longman.

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor, & T. A.Upton (Eds.), *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123-145). Amsterdam & Atlanta: Rodopi.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, *1*(1), 7-24.

Granger, S., Gilquin, G., & Meunier, F. (2016). *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: Benjamins.

Gravano, A., & Hirschberg, J. (2009). Backchannel-inviting cues in task-oriented dialogue. In *Proceedings of 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *September 6-10 2009*, 1019-1022. Brighton, United Kingdom. <https://www.isca-speech.org/archive/interspeech_2009/i09_1019.html>. (12 September 2018).

Grice, H. P. (1975). Logic and Conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and Semantics 3*: *Speech Acts* (pp. 41-58). New York : Academic Press.

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, *191*, 267-298.

Grice, M., & Baumann, S. (2007). An introduction to intonation - functions and models. In J. Trouvain, & U. Gut (Eds.), *Non-Native Prosody - phonetic description and teaching practice* (pp. 25-51). Berlin: Mouton de Gruyter.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental - metrical phonology. In S. - A. Jun (Ed.), *Prosodic Typology* (pp. 55-83). Oxford: Oxford University Press.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Consistency in Transcription and Labelling of German Intonation with GToBI. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP), October 3-6 1996*, 1716-1719. Philadelphia, PA, USA. <https://www.isca-speech.org/archive/icslp_1996/i96_1716.html>. (4 August 2017).

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90-107.

Gries, S. Th. (2013). *Statistics for Linguistics with R. A Practical Introduction*. De Gruyter Mouton.

Grieve , A. (2010). 'Aber ganz ehrlich': differences in episodic structure, apologies and truth-orientation in German and Australian workplace telephone discourse. *Journal of Pragmatics*, *42*(1), 190-219.

Grosjean, P., & Ibanez, F. (2018). Pastecs: Package for analysis of space-time ecological series. R package version 1.3.21. https://CRAN.R-project.org/package=pastecs

Grosser, W. (1982). *Intonatorische Inteferenzerscheinungen Deutsch - English*. Wien: Verband der Wissenschaftlichen Gesellschaften Österreichs.

Grosser, W. (1989). *Akzentuierung und Intonation im englischen L2-Erwerb österreichischer Lerner*. Salzburg: Salzburger Studien zur Anglistik und Amerikanistik 9.

Grosser, W. (1993). Aspects of intonational L2 acquisition. In. B. Kettemann, & W. Wieden (Eds.), *Current nissues in European second language acquisition research* (pp. 81-94). Tübingen: Gunter Narr Verlag.

Grosser, W. (1997). On the acquisiton of tonal and accentual features of English by Austrian learners. In A. James, & J. Leather (Eds.), *Second Language Speech: Structure and Process* (pp. 207-224). Berlin: Walter de Gruyter.

Grosvald, M. (2009). Inter-speaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, *37*(2), 173-188.

Grover, C., Jamieson, D. G., & Dobrovolsky, M. B. (1987). Intonation in English, French, and German: Perception and production. *Language and Speech, 30*(3), 277-295.

Gumperz, J. J. (1982). *Discourse strategies*. Studies in Interactional Sociolinguistis 1. Cambridge: Cambridge University Press.

Gunlogson, C. A. (2001). True to form: Rising and falling declaratives as questions in English. PhD. University of California at Santa Cruz, Santa Cruz, CA.

Gussenhoven, C. (1984). On the grammar and semantics of sentence accents. Dordrecht: Foris.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In *Proceedings of Speech Prosody*, *April 11-13 2002,* B. Bel, & I. Marlin (Eds.), 47-57. Aix-en-Provence, France: Laboratoire Parole et Langage. <https://www.isca-speech.org/archive/sp2002/sp02_047.html>. (03 May 2017).

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Gut, U. (2000). On the acquisition of rhythmic structure. In *Proceedings of New Sounds* 2000, *September 2000*, 148-154. Amsterdam, The Netherlands.

Gut, U. (2007). Sprachkorpora im Phonetikunterricht. *Zeitschrift für Interkulturellen Fremd-sprachenunterricht*, *12*(2), 1-21.

Gut, U. (2009). *Non-Native Speech: A Corpus-Based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt: Peter Lang.

Gut, U. (2010). The LeaP corpus. A phonetically annotated corpus of non-native speech. <http://www.philhist.uni-augsburg.de/de/lehrstuehle/anglistik/angewandte_sprachwissenschaft/workshop/pdfs/LeapCorpus_Manual.pdf>.

Gut, U. (2012). The LeaP corpus: A multilingual corpus of spoken learner German and learner English. In T. Schmidt, & K. Wörner (Eds.), *Hamburg studies on multilingualism: Vol. 14. Multilingual corpora and multilingual corpus analysis* (pp. 3-23). Amsterdam, NL: John Benjamins.

Gut, U. (2014). *Corpus Phonology and Second Language Acquisition*. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford Handbook of Corpus Phonology* (pp. 286-301). Oxford: Oxford University Press.

Gut, U. (2017). Phonological development in different learning contexts - a corpus-based study. *International Journal of Learner Corpus Research* 3(2). 196-222.

Gut, U., Trouvain, J., & Barry, W. J. (2007). Bridging research on phonetic descriptions with knowledge from teaching practice - The case of prosody in non-native speech. In J. Trouvain, & U. Gut (Eds.), *Non-Native Prosody: Phonetic Description and Teaching Practice. Trends in Linguistics. Studies and Monographs 186* (pp. 3-24). Berlin: De Gruyter Mouton.

Gutiérrez Díez, F. (1982). Bases para una comparación de la prosodia del catalán, el castellano y el ingles. PhD. University de Barcelona.

Gutiérrez Díez, F. (1983). Aspectos lingüísticos de la segmentación del tone en ingles, castellano y catalán. In *Actas del 1 Congreso Nacional de Lingüística Aplicada*, 179-181. Madrid: SGEL, Universidad de Murcia.

Gutiérrez Díez, F. (1995). *La function demarcative de la entonación en ingles castellano y catalán*. Murcia: Servicio de Publicaciones de la Universidad de Murcia.

Gutiérrez Díez, F. (2005). Intonation focus in the interlanguage of a group of Spanish learners of English. *Revista Alicantina de Estudios Ingleses*, *18*, 127-147.

Gutiérrez Díez, F. (2008). Contrastive intonation and error analysis. Tonality and tonicity in the interlanguage of a group of Spanish learners of English. In M. A. Gómez, J. Lachlan Mackenzie, & E. González (Eds.), *Languages and cultures in contrast and comparison* (pp. 327-354). Amsterdam: Johnson Benjamins.

Gutiérrez Díez, F. (2012). Trouble spots in the learning of English intonation by Spanish speakers. Tonality and tonicity. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 219-230). Springer Science + Business Media Dodrecht.

Guy, G., & Vonwiller, J. (1989). The high rising tone in Australian English. In P. Collins, & D. Blair (Eds.), *Australian English* (pp. 21-34). University of Queensland Press.

Guy, G., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I. (1986). An intonational change in Australian English. *Language in Society*, *15*(1), 23-51.

Halliday, M. A. K. (1963). Intonation in English grammar. *Transactions of the Philological Society*, *62*(1), 143-169

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. (1970). *A course in spoken English: Intonation*. London: Oxford University Press.

Halpern, D. F. (2012). *Sex differences in cognitive abilities.* (4th ed.). New York, NY, US: Psychology Press.

Han, Z. H. (2009). Interlanguage and fossilization: Towards an analytic model. In V. Cook, & L. Wei (Eds.), *Contemporary Applied Linguistics 1. Language Teaching and Learning* (pp. 137-162). Continuum International Publishing.

Han, Z. H. (2013). Forty years later: Updating the Fossilization Hypothesis. *Language Teaching*, *46*(2), 133-171.

Hanley, T. D., & Snidecor, J.C. (1967). Some acoustic similarities among languages. *Phonetica*, *17*, 141-148.

Hanley, T. D., Snidecor, J.C., & Ringel, R. (1966). Some acoustic differences among languages, *Phonetica*, *14*, 97-107.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, *8*, 34-52.

Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal*, *22*(2), 175-190.

Hartshorne, J. Joshua, K., Tenenbaum, B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263-277.

Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks?: The implied communality deficit. *Journal of Applied Psychology, 92*(1), 81-92.

Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U., & Curnick, L. (2012). *The English pronunciation teaching in Europe survey: Selected results. Research in Language*, *10*(1), 5-27.

Henriksen, N. C., Geeslin, K. L., & Willis, E. W. (2010). The development of L2 Spanish intonation during a study abroad immersion program in León, Spain. *Studies in Hispanic and Lusophone Linguistics, 3*(1), 113-162.

Henton, C. G. (1989). Fact and fiction in the description of female and male speech. *Language and Communication*, *9*, 299-311.

Henton, C. G. (1995). Pitch dynamism in female and male speech. *Language and Communication*, *15*, 43-61.

Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *International Review of Applied Linguistics in Language Teaching, 33*(3), 251-265.

Hincks, R. (2004). Processing the prosody of oral presentations. In *Proceedings of InSTIL/ICALL 2004 Symposium on Computer Assisted Learning*, *June 17-19 2004*, 63-66. Venice, Italy. <https://www.isca-speech.org/archive_open/icall2004/iic4_016.html>. (1 November 2017).

Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology*, *13*(3): 32-50.

Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In M. Horne (Ed.), *Prosody: Theory and Experiment. Studies presented to Gösta Bruce* (pp. 335-350). Dordrecht: Kluwer Academic Publishers.

Hirschfeld, U., & Trouvain, J. (2007). Teaching prosody in German as foreign language. In J. Trouvain, & U. Gut (Eds.), *Non-Native Prosody - phonetic description and teaching practice* (pp. 171-187). Berlin: Mouton de Gruyter.

Hirson, A., French, P., & Howard, D. (1995). Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In J. Windsor Lewis (Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor* (pp. 230-240). London: Routledge.

Hirst, D. Auran, C., & Bouzon, C. (2009). *The Aix-MARSEC database*. 2002-2004. Tech. Report, Equipe Prosodie et Représentation Formelle du Langage, Laboratoire CNRS UMR 6057 Parole et Langage, Université de Provence, Aix-en-Provence.

Hirst, D., & di Cristo, A. (1998). A survey of intonation systems. D. Hirst, & A. di Cristo (Eds.), *Intonation Systems* (pp. 1-44). Cambridge: Cambridge University Press.

Hoff, M. A. (2014). *Ethnic Identity and Accent: Exploring Phonological Acquistion for International Students from China*. MA thesis, Bowling Green State University.

Hollien, H. (1990). *Acoustics of Crime: The New Science of Forensic Phonetics.* Springer Science: New York.

Hollien, H., & Paul, P. (1969). A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls. *Language and Speech, 12*, 119-124.

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronological age in males. *Journal of Speech and Hearing Research, 15*, 155-159.

Hollien, H., Hollien, P. A., & De Jong, G. (1997). Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America*, *102*, 2984-2992.

Holmes, J. (1995). *Women, Men and Politness*. London: Longman.

Hoque, M. E., Courgeon, M., Martin, J.- C., Mutlu, B., & Picard, R. W. (2013). Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing (UBICOMP)*, *September 8-12 2013*, 697-706. Zürich, Switzerland. <https://dl.acm.org/doi/pdf/10.1145/2493432.2493502>. (23 November 2017).

Horri, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of Speech and Hearing Research*, *18*, 192-201.

Horvath, B. (1985). *Variation in Australian English: The Sociolects of Sydney.* Cambridge University Press.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. 2nd edition. New York: Wiley.

Hothorn, T., & Zeileis, A. (2015). Partykit: A Modular Toolkit for Recursive Partytioning in R. Journal of Machine Learning Research, 16, 3905-3909. URL <http://jmlr.org/papers/v16/hothorn15a.html> (12 January 2016).

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651-674.

House, J. (2006). Constructing a context with intonation. *Journal of Pragmatics*, *38*(10), 1542-1558.

Howell, D. C. (2013). *Statistical methods for Psychology*. 8th edition (international edition). Cengage Learning.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech. *Speech Communication, 10*, 163-169.

Hualde, J. I. (2002). Intonation in Spanish and the other Ibero-Romance languages: Overview and status quaestionis. In *Romance Phonology and Variation: Selected Papers from the 30th Linguistic Symposium on Romance Languages, February 2000,* C. R. Wiltshire, & J. Camps (Eds.), 101-115. *Gainesville, Florida,* Amsterdam: John Benjamins.

Hualde, J. I. (2003). Romance Intonation from a Comparative and Diachronic Perspective. In *Contemporary approaches to romance linguistics*. *Selected papers from the 33rd linguistic symposium on Romance languages (LSRL), April 2003*, J. Auger, J. C. Clements, & B. Vance (Eds.), 217-237.Bloomington, Indiana.

Hualde, J. I., & Prieto, P. (2016). Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1-25. http://doi.org/10.5334/labphon.11

Huang, L.- F. (2019). The construction of spoken English learner corpus from several L1s across CEFR levels. The Ministry of Science and Technology, Republic of China. Grant number: MOST 107-2410-H-012-001. Grant amount: USD 23000, Republic of China Naval Academy, Duration: 1 Aug 2018 - 30 September 2019 (14 months).

Huang, L.- F., Kubelec, S., Keng, N., & Hsu, L.-H. (2018). Evaluating CEFR rater performance through the analysis of spoken learner corpora. *Language Testing in Asia*, *8*(14), 1-17.

Hudson, T., De Jong, G., McDougall, K., Harrison, P., & Nolan, F. (2007). F0 statistics for 100 young male speakers of standard southern British English. In *16th International Congress of Phonetic Sciences (ICPhS), August 6-10 2007*, 1809-1812. Saarbrücken, Germany.

Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquistion*, *11*, 241-255.

Hurford, J. R. (1991). The evolution of the critical period for language acquisition. *Cognition*, *40*, 159-201.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Pychological Bulletin*, *104*, 53-69.

Hyde, M. (1994). The importance of language awareness in the phonology component of a teacher training programme. *Language Awareness*, *3*(3-4), 141-150.

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a second language? All, some, or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica, 54,* 150-166.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty, & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 539-588). Malden, MA: Blackwell.

Ibarraran, A., Lasagabaster, D., & Sierra, J. M. (2008). Multilingualism and language attitudes: Local versus immigrant students' perceptions. *Language Awareness*, *17*(4), 326-41.

Idsardi, W. (1992). The computation of prosody. Unpublished MIT thesis.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73-98.

Ishikawa, S. (2015). Lexical development in L2 English learners' speeches and writings. *Procedia*, *198*, 202-210.

Jacewicz, E., Fox, R. A., & O’Neill, C. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*, 233-256.

Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, *128*, 839-850.

Jackson, C. N., & O’Brien, M. G. (2011). The interaction between prosody and meaning in second language speech production. *Die Unterrichtspraxis*, *44*(1), 1-11.

Jakšič, J., & Šturm, P. (2017). Accents of English at Czech schools: Students’ attitudes and recognition skills. *Research in Language*, *15*(4), 353-369.

James, E., Mahut, C., & Latkiewitcz, G. (1989). The investigation of an apparently new intonation pattern in Toronto English. *Information Communication*, *10*, 11-17.

Jansen, W., Gregory, M. L., & Brenier, J. M. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. In *ITRW on Prosody in Speech Recognition and Understanding*, *October 22-24 2001*, M. Pitcher Inn (Ed.), paper 14. Red Bank, NJ, USA. <https://www.isca-speech.org/archive_open/prosody_2001/prsr_014.html>. (01 April 2017).

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, *50*(2), 245-309.

Jarvis, S., & Pavlenko, A. (2009). *Crosslinguistic Influence in Language and Cognition.* Abingdon: Routledge.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, *23*(1), 83-103.

Jessen, M., Köster, O., & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, *12*, 174-213.

Jiang, M., Green, R. J., Henley, T. B., & Masten, W. G. (2009). Acculturation in relation to the acquisition of a second language. *Journal of Multilingual and Multicultural Development*, *30*, 481-492.

Jilka, M. (2000). *The Contribution of Intonation to the Perception of Foreign Accent*. PhD thesis, Institute of Natural Language Processing, University of Stuttgart.

Jilka, M. (2007). Different manifestations and perceptions of foreign accent in intonation. In J. Trouvain, & U. Gut (Eds.), *Non-Native Prosody - phonetic description and teaching practice* (pp. 77-96). Berlin: Mouton de Gruyter.

Jiménez Vilches, R. (2017). Analysing four prosodic forms to convey the pragmatic function of dominance and control in L2 monologic discourse. *EPiC Series in Language and Linguistics*, *2*, 199-213.

Jiránková, L., Gráf, T., & Kvítková, A. (2019). On the relation between L1 and L2 speech rate. In A. Abel, A. Glaznieks, V. Lyding, & L. Nicolas (Eds.), *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference. Corpora and Language in Use - Proceedings 5*, (pp. 19-41). Louvain-la-Neuve: Presses universitaires de Louvain.

Jiřelová, K. (2018). Language transfer and fluency. Diploma thesis. Charles University, Prague.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psycho1ogy*, *21*, 60-99.

Jun, S.- A., & Fletcher, J. (2014). Methodology of Studying Intonation: From Data Collection to Data Analysis. In S. - A. Jun (Eds.), *Prosodic Typology II: The New Development in the Phonology of Intonation and Phrasing* (pp. 493-519). Oxford: Oxford University Press.

Jun, S.- A., & Oh, M. (2000). Acquisition of second language intonation. *Journal of the Acoustical Society of America*, *107*(5), 2802-2803.

Kachru, B. B. (1994). Englishization and contact linguistics. *World Englishes*, *13*(2), 135-154.

Kachru, B. B. (1996). World Englishes: Agony and ecstasy. *Journal of Aesthetic Education*, *30*(2), 135-155.

Kallman, H. J., & Massaro, D. W. (1983). Backward masking, the suffix effect, and preperceptual storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(2), 312-327.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System, 38*, 301-315.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554-566.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differencs in L1 production on L2 pronunciation. *Frontiers in Psychology*, *5*, 1-17.

Kasper, G. (1981). *Pragmatische Aspekte in der Interimsprache*. Tübingen: Gunter Narr.

Kehrein, R. (2002). *Prosodie und Emotionen*. Reihe Germanistische Linguistik 231. De Gruyter.

Kelm, O. R. (1995). Acoustic measurements of Spanish and English pitch contours: Native and non-native speakers. *Hispanic Linguistics*, *6*(7), 435-448.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, *64*(3), 459-489.

Keseling, G. (1992). Pause and intonation contours in written and oral discourse. In D. Stein (Ed.), *Cooperating with Written Texts. The Pragmatics and Comprehension of Written Texts* (pp. 31-66). Berlin: Mouton de Gruyter.

Kim, M. (2012). Phonetic accommodation after auditory exposure to native and nonnative speech. PhD thesis, Northwestern University, Evanston, IL.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125-156.

Kimura, D. (1999). *Sex and cognition.* Cambridge, MA: MIT Press.

Kingdon, R. (1958). *The Groundwork of English Intonation*. London: Longman.

Kjellmer, G. (2003). Hesitation. In Defence of Er and Erm. *English Studies*, *84*(2), 171-198.

Klein, W., & Perdue, C. (1997). The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, *13*, 301-347.

Klewitz, G., & Couper-Kuhlen, E. (1999). Quote-unquote. The role of prosody in the contextualization of reported speech sequences. *Pragmatics*, *9*(4), 459-485.

Knoll, M. A., & Costall, A. (2015). Characterising F(0) contour shape in infant - and foreigner-directed speech. *Speech Communication*, *66*, 231-243.

Knoll, M. A., & Uther, M. (2004). Motherese and Chinese: Evidence for acoustic changes in speech directed at infants and foreigners. *Journal of the Acoustical Society of America, 116*(4), 2522.

Knoll, M. A., Uther, M., MacLeod, N., O'Neill, M., & Walsh, S. (2006). Emotional, linguistic or just cute? The function of pitch contours ininfant- and foreigner-directed speech. In *Proceedings of Speech Prosody 3rd International Conference*, *May 2-5 2006*, 165-168. Dresden, Germany. <https://www.isca-speech.org/archive/sp2006/sp06_196.html>. (11 September 2017).

Knowles, G. (1973). *Scouse: The urban dialect of Liverpool*. PhD thesis. University of Leeds Leeds.

Kramsch, C. (1998). The privilege of the intercultural speaker. In M. Byram, & M. Fleming (Eds.), *Language Learning in Intercultural Perspective: Approaches through Drama and Ethnography* (pp. 16-31). Cambridge: Cambridge University Press.

Krippendorff, K. (2004). *Content Analysis: An introduction to its methodology*. 2nd edition. Sage Publications.

Künzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, *46*, 117-125.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786

Labov, W. (1966). *The social stratification of English in New York City*.Washington: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language*, *59*, 721-59.

Ladd, D. R. (1988). Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, *84*, 530-44.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Ladd, D. R. (2008). Intonational Phonology. 2nd edition. Cambridge: Cambridge University Press.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, *106*, 1543-1554.

Ladd, R. D., Silverman, K. E. A., Tolkmitt, F., Bergmann, G., & Scherer, K. R. (1985). Evidence for the independent function of intonation contour type, voice quality, and f0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*, 435-444.

Lado, R. (1957). *Linguistics Across Cultures:applied linguistics for language teachers*. University of Michigan Press.

Lakoff, R. (1975). *Language and women's place.* New York: Harper & Row.

Lasagabaster, D. (2005). Attitudes towards Basque, Spanish and English: an analysis of the most influential variables. *Journal of Multilingual and Multicultural Development*, *26*(4), 296-316.

Lasagabaster, D., & Sierra, J. M. (2005). The nativeness factor: an analysis of students' preferences. *ITL International Journal of Applied Linguistics*, *148*, 21-43.

Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech Language and Hearing Research*, *42*(5), 1148-56.

Le Bruyn, B., & Paquot, M. (Eds.). (2020), *Second Language Acquisition and Learner Corpora*. Cambridge: Cambridge University Press.

Lee, J. (2007). The inside of the L2 lexicon and semantic overgeneralization. Unpublished manuscript, Michigan State University.

Lee, W. R. (1958). *English Intonation: A New Approach*. Amsterdam: North Holland.

Leeman, A. (2009). Intonational variation in Swiss German. In *Language Variation - European Perspectives II: Selected Papers from the 4th International Conference on Language Variation in Europe (ICLaVE 4), June 2007*, S. Tsiptakou, M. Karylomenou, & P. Y. Pavlou (Eds.), 135-143. Nicosia, Amsterdam: John Benjamins.

Lengeris, A. (2012). Teaching and learning English prosody: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 25-40). Dordrecht: Springer.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning*, *40*(3), 387-417.

Lentz, T. O., & Chen, A. (2015). Unbalanced adult production and perception of prosody. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. < https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0972.pdf>. (05 June 2017).

Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, *39*, 397-413.

Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation*. Cambridge MA: MIT Press.

Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, *33*(1), 37-64.

Levis, J. (2005). Comparing apples and oranges? Pedagogical approaches to intonation in British and American English. In K. Dziubalska-Kolaczyk, & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (pp. 339-366). Bern: Peter Lang.

Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press.

Levis, J. (2019). Cinderella no more: Leaving victimhood behind. *Speak Out*, *60*(1), 1-7.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*(4), 505-524.

Levis, J., & Sonsaat, S. (2019). Quality Teacher Education For Pronunciation Teaching in L2 Classrooms. In J. de Dios Martínez Agudo (Ed.), *Quality in TESOL and Teacher Education. From a Results Culture Towards a Quality Culture* (pp. 213-222). New York and London: Routledge.

Levitan, R. (2014). *Acoustic-prosodic entrainment in human-human and human-computer dialogue*. PhD thesis. Columbia University.

Levitan, R., & Hirschberg, J. (2011). Measuring acousticprosodic entrainment with respect to multiple levels and dimensions. In *Proceedings of Interspeech*, *12th Annual Conference of the International Speech Communication Association (INTERSPEECH), August 27-31 2011*, P. Cosi, R. De Mori, G. Di Fabbrizio, & R. Pieraccini, 3081-3084. Firenze, Italy. <https://www.isca-speech.org/archive/interspeech_2011/i11_3081.html>. (05 May 2018).

Levon, E. (2016). Gender, interaction and intonational variation: The discourse functions of high rising terminals in London. *Journal of Sociolinguistics*, *20*(2), 133-163.

Levon, E. (2018). Same difference: the phonetic shape of high rising terminals in London. *English Language and Linguistics*, *24*(1), 1-25.

Levon, E., & Ye, Y. (2020). Language, Indexicality and Gender Ideologies: Contextual Effects on the Perceived Credibility of Women. *Gender and Language*, 14(2), 123-151.

Levshina, N. (2015). *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins Publishing.

Lewandowski, E., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *144*(2), 620-633.

Lewandowski, N. (2012). Talent in nonnative phonetic convergence. PhD thesis. University Stuttgart.

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, *4*(18), 1-19.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, *54*(2), 193-226.

Liberman, M., & Prince, A. (1977). On Stress and Linguistics Rhythm. *Linguistic Inquiry*, *8*(2), 249-336.

Lightbown, P. M., & Spada, N. (2006). *How Languages are Learned*. Oxford: Oxford University Press.

Lightbown, P. M., & Spada, N. (2013). *How Languages are Learned*. Oxford: Oxford University Press.

Lin, P. M. S. (2011). Perception and production of prosodically varying inter-gestural timing in American English laterals. PhD thesis. University of Michigan, Ann Arbor.

Lin, P. M. S. (2012a). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics, 33*(3), 342-347.

Lin, P. M. S. (2012b). More than music to our ears: The value of the phonological interface in a comprehensive understanding of vocabulary acquisition and knowledge. In *Multilingual Teory and Practice in Applied Linguistics: Proceedings of the 45th Annual*

Meeting of the British Association for Applied Linguistics, September 6-8 2012, A. N. Archibald (Ed.), 155-158. University of Southampton. London: Scitsiugnil Press.

Lin, P. M. S. (2013). The prosody of formulaic expressions in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics*, *18*(4), 561-588.

Lin, P. M. S. (2018). The prosody of formulaic sequences: A corpus and discourse approach. In W. Teubert, & M. Mahlberg (Eds.), *Research in Corpus and Discourse*. Bloomsbury.

Lindsey, G. (2019). *English after RP: Standard British Pronunciation Today*. Palgrave Macmillan.

Linneman, T. J. (2013). Gender in Jeopardy!: Intonation variation on a television game show. *Gender & Society*, 27(1), 82-105.

Lintunen, P., Peltonen, P,.& Webb, J. (2015). Tone units as indicators of L2 fluency development: evidence from native and learner English. In J. A. Mompean, & J. Fouz-González (Eds.), *Investigating English Pronunciation: Trends and Directions* (pp. 196-218). Basingstoke: Palgrave Macmillan.

Liscombe, J., Hirschberg, J., & Venditti, J. J. (2005). Detecting Certainness in Spoken Tutorial Dialogues. In *Proceedings of 9th European Conference on Speech Communication and Technology Eurospeech (INTERSPEECH)*, *September 4-8 2005*, 1837-1840. Lisbon, Portugal. <https://www.isca-speech.org/archive/interspeech_2005/i05_1837.html>. (01 June 2018).

Litman, D., & Forbes, K. (2003). Recognizing Emotions from Student Speech in Tutoring Dialogues. In *Automatic Speech Recognition and Understanding, 2003. ASRU '03. 2003 IEEE Workshop*, *November 30 - December 4 2003*. St Thomas, VI, USA. < https://ieeexplore.ieee.org/document/1318398>. (02 April 2017).

Litman, D., Rotaru, M., & Nicholas, G. (2009). Classifying turn-level uncertainty using word-level prosody. In *Proceedings of 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *September 6-10 2009*, 2003-2006. Brighton, United Kingdom. <https://www.isca-speech.org/archive/interspeech_2009/i09_2003.html>. (02 August 2018).

Lleó, C., & Vogel, I. (2004). Learning new segments and reducing domains in German L2 phonology: the role of the prosodic hierarchy. *International Journal of Bilingualism*, *8*, 79-104.

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4,* 126-141.

Long, M H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-85.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), Handbook of research on Language Acquisition: Second language acquisition 2 (pp. 413-68). New York: Academic Press.

Long, M. H. (2003). Stabilization and fossilization in second language interlanguage development. In C. Doughty, & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 487-529). Oxford: Blackwell.

López Rúa, P. (2006). The sex variable in foreign language learning: an integrative approach. *Porta Linguarum*, *6*, 99-114.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into thepitch correlates of English and Japanese politeness formulae. *Language and Speech*, *24*(1), 71-89.

Loveday, L. J. (1982). *The Sociolinguistics of Learning and Using a Non-Native Language*. Oxford: Pergamon Press.

Lowry, O. (2002). The stylistic variation of nuclear patterns in Belfast English. *Journal of the International Phonetic Association*, *32*(1), 33-42.

Lyons, J. (1981). *Language and Linguistics. An Introduction*. Cambridge: Cambridge University Press.

Macaulay, R. K. S. (1978). The myth of female superiority in language. *Journal of Child Language*, *5*(2), 353-363.

Markham, D. (1997). *Phonetic Imitation, Accent, and the Learner*. Lund: Lund University Press.

Marzec-Stawiarska, M. (2014). Gender differences in foreign language speaking-in-class anxiety. *Linguistica Silesiana*, *35*, 417-434.

Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, *32*(1), 21-36.

Mauranen, A. (2006). Signalling and preventing misunderstanding in English as a lingua france communication. *International Journal of the Sociology of Language*, 177, 123-150.

Mayer, J. (1995). *Transcription of German intonation - the Stuttgart System*. Technical report, University of Stuttgart.

McGregor, J. (2006). *High rising tunes in Australian English.* Unpublished PhD thesis. Macquarie University.

McKirnan, D. J., & Hamayan, E. V. (1984). Speech norms and attitudestoward outgroup members: A test of a model in a bicultural context. J*ournal of Language and Social Psychology*, *3*(1), 21-38.

McLemore, C. A. (1990). The interpretation of L*H in English. *Texas Linguistic Forum 32: Discourse, University of Texas, Austin*, 127-147.

McLemore, C. A. (1991). *The Pragmatic Interpretation of English Intonation: Sorority Speech*. PhD thesis. University of Texas at Austin.

Méndez Seijas, J. (2018). L2 Spanish intonation in a short-term study abroad program. In C. Sanz, & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp. 86-100). New York and London: Routledge.

Mennen, I. (1998). Can language learners ever acquire the intonation of a second language? In Proceedings *ETRW on Speech Technology in Language Learning (STiLL), May 25-27 1998,* 17-20. Marholmen, Sweden. <https://www.isca-speech.org/archive_open /still98/stl8_017.html>. (15 August 2016).

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of phonetics*, *32*, 543-563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. J. Trouvain, & U. Gut (Eds.), *Non-Native Prosody - phonetic description and teaching practice* (pp. 53-76). Berlin: Mouton de Gruyter.

Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171-188). Berlin: Springer.

Mennen, I., & de Leeuw, E. (2014). Beyond Segments. Prosody in SLA. *Studies in Second Langauge Acquistion*, *36*, 183-194.

Mennen, I., Chen, A., & Karlsson, F. (2010). Characterising the internal structure of learner intonation and its development over time. In *Proceedings of New Sounds 2010: The sixth international symposium on the acquisition of second language speech*, *May 1-3 2010*, K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), 319-324. University of Poznań, Poland. <http://ifa.amu.edu.pl/newsounds/files/proceedings/proceedings_quotable_version.pdf>. (13 September 2015).

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second Language Acquistion of pitch range in German learners of English. *Studies in Second Language Acquistion*, *36*, 303-329.

Mennen, I., Schaeffler, F., & Docherty, G. (2007). Pitching it differently: A comparison of the pitch ranges of German and English speakers. In *16th International Congress of Phonetic Sciences (ICPhS), August 6-10 2007*, 1769-1972. Saarbrücken, Germany.

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: a comparison of English and German. *Journal of the Acoustical Society of America*, *131*(3), 2249-2260.

Michalsky, J., & Schoormann, H. (2017). Pitch convergence as an effect of perceived attractiveness and likability. In *Proceedings of 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *August 20-24 2017*, 2253-2256. Stockholm, Sweden. <https://www.isca-speech.org/archive/Interspeech_2017/abstracts/1520.html>. (25 August 2017).

Michalsky, J., & Schoormann, H. (2018). Opposites attract! Pitch divergence at turn breaks as cause and effect perceived attractiveness. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 265-268. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/70.html>. (18 June 2018).

Michas, I. C., & Berry, D. C. (1994). Implicit and explicit processes in a second language learning task. *European Journal of Cognitive Psychology*, *6*(4), 357-381.

Milde, J.-T., & Gut, U. (2002). A prosodic corpus of non-native speech. In *Proceedings of Speech Prosody*, B. Bel, & I. Marlien (Eds.), 503-506. Aix-en-Provence, France: Laboratoire Parole et Langage. <https://www.isca-speech.org/archive/sp2002/sp02_503.html>. (1 May 2017).

Miles, S. A., Miranda, R. A., & Ullman, M. T. (2016). Sex differences in music: A female advantage at recognizing familiar melodies. *Frontiers in Psychology*, *7*, 1-12.

Miller, J. E., & Weinert, R. (2009). *Spontaneous Spoken Language: Syntax and Discourse*. Oxford; New York: Oxford University Press.

Missaglia, F. (1999). Contrastive prosody in SLA - an empirical study with adult Italian learners of German. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99)*, *Vol. 1, August 1-7 1999,* J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), 551-554. San Francisco, United States. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0551.pdf>. (1 June 2016).

Möbius, B. (1993). *Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverläufen*. Tübingen : Niemeyer.

Mok, P., Yin, Y., Setter, J., & Nayan, N. M. (2016). Assessing knowledge of English intonation patterns by L2 speakers. In *Proceedings of Speech Prosody, May 31 - June 3 2016*, J. Barnes, A. Brugos, S. Shattuck-Hufnagel, & N. Veilleux (Eds.), 1-5. Boston, USA. <https://www.isca-speech.org/archive/SpeechProsody_2016/ pdfs/49.pdf>. (2 August 2017).

Mompeán González, J. A. (2004). Options and criteria for the choice of an English pronunciation model in Spain. In J. Anderson, J. M. Oro, & J. Varela Zapata (Eds.), *Linguistic perspectives from the classroom: language teaching in a multicultural Europe* (pp. 1043-1059). Universidade de Santiago de Compostela.

Monroy, R., & Hernández-Campoy, J. M. (2015). Murcian Spanish. *Journal of the International Phonetic Association*, *45*(2), 229-240.

Moore, B., & Glasberg, B. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, *74*(3), 750-753.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81-108.

Moyer, A. (2004). *Age, accent, and experience in second language acquisition: an integrated approach to critical period inquiry*. Clevedon, UK: Multilingual Matters Ltd.

Moyer, A. (2007). Do language attitudes determine accent? A study of bilinguals in the USA. *Journal of Multilingual and Mulitcultural Development*, *28*, 1-17.

Moyer, A. (2011). An investigation of experience in L2 phonology: Does Quality Matter More than Quantity? *Canadian Modern Language Review*, *67*, 191-216.

Moyer, A. (2013). *Foreign accent: the phonology of non-native speech*. Cambridge: Cambridge University Press.

Moyer, A. (2016). The puzzle of gender effects in L2 phonology. *Journal of Second Language Pronunciation*, *2*(1), 8-28.

Moyer, A. (2017). Autonomy in second language phonology: Choice vs. limits. *Language Teaching*, *50*(3), 395-411.

Moyer, A. (2018). An Advantage for Age? Self-Concept and Self-Regulation as Teachable Foundations in Second Language Accent. *The CATESOL Journal*, *30*(1), 1-18.

Mukherjee, J. (2001). *Form and Function of Parasyntactic Presentation Structures: A Corpus-based Study of Talk Units in Spoken English*. Amsterdam: Rodopi.

Mukherjee, J. (2002). *Korpuslinguistik und Englishunterricht: Eine Einführung*. Frankfurt am Main: Peter Lang.

Mukherjee, J. (2005). The native speaker is alive and kicking: Linguistic and language-pedagogical perspectives. *Anglistik*, *16*(2), 7-23.

Mukherjee, J., & Rohrbach, J.-M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. B. Kettemann, & G. Marko (Eds.), *Planning, Gluing and Painting Corpora* (pp. 205-232). Frankfurt am Main: Peter Lang.

Müller, S. (2005). *Discourse Markers in Native and Non-Native English Discourse*. John Benjamins, Amsterdam/Philadelphia.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners, *Language Learning*, *49*(1), 285-310.

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, *21*(4), 373-391.

Nakajima, S., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, *50*, 197-210.

Namy, L. L., Nygaard L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422-432.

Nance, C., Kirkham, S., & Groarke, E. (2015). Intonational Variation in Liverpool English. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0633.pdf>. (04 Januar 2016).

Nance, C., Kirkham, S., & Groarke, E. (2018). Studying intonation in varieties of English: Gender and individual variation in Liverpool. In N. Braber, & S. Jansen (Eds.), *Sociolinguistcs in England* (pp. 275-295). London: Palgrave Macmillan.

National Institute for Education. (2017). *Rámcový vzdělávací program pro základní vzdělávání*. <http://www.nuv.cz/uploads/RVP_ZV_2017_verze_cerven.pdf>. (15 November 2019).

Nava, E. (2008). Prosody in L2 Acquisition. In *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)*, *May 18-20 2007*, R. Slabakova, J. Rothman, P. Kempchinksy, & E. Gavruseva (Eds.), 155-164. University of Iowa. Somerville, MA: Cascadilla Proceedings Project.

Nava, E., & Zubizarreta, E. (2009). Order of L2 acquisition of prosodic prominence patterns: Evidence from L1 Spanish/L2 English speech. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*, *September 4-6 2008*, J. Crawford, K. Otaki, & M. Takahashi (Eds.), 175-187. University of Connecticut in Storrs. Somerville, MA: Cascadilla Press.

Neary-Sundquist, C. A. (2008). *The role of task type and proficiency level in second language speech production.* PhD thesis. Purdue University.

Nevalainen, T. (1992). Intonation and discourse type. *Text*, *12*, 397-427.

New South Voices Collection, J. Murrey Atkins Library, The University of North Carolina at Charlotte.

Nibert, H. J. (2000). *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation*. PhD thesis. University of Illinois, Champaign.

Nibert, H. J. (2005). The acquisition of the phrase accent by intermediate and advanced adult learners of Spanish as a second language. In *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, D. Eddington (Ed.), 108-122. Somerville, MA: Cascadilla Proceedings Project.

Nibert, H. J. (2006). The acquisition of the phrase accent by beginning adult learners of Spanish as a second language. In *Selected Proceedings of the* 199 *2nd Laboratory*

Approaches to Spanish Phonetics and Phonology, M. Díaz-Campos (Ed.), 131-148. Somerville, MA: Cascadilla Proceedings Project.

Niebuhr, O. (2015). Gender Differences in the Prosody of German Questions. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0051.pdf>. (05 February 2016).

Niebuhr, O., & Ward, N. G. (2018). Challenges in studying prosody and its pragmatic functions: Introduction to JIPA special issue. *Journal of the International Phonetic Association*, *48*(1), 1-8.

Nilsenová, M. (2006). *Rises and Falls: Studies in the Semantics and Pragmatics of Intonation*. PhD thesis. University of Amsterdam.

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences*, *August 3-9 2003*, M.J. Solé, D. Recasens, & J. Romero (Eds.), 771-774. Barcelona: Universitat Autònoma de Barcelona.

Nowacka, M. (2012). Questionnaire-based pronunciation studies: Italian, Spanish and Polish students' views on their English pronunciation. *Research in Language*, *10*(1), 43-61.

O'Barr, W., & Atkins, B. (1998 [1980]). 'Women's language' or 'powerless language'? In Gita DasBender (Ed.), *Language: a Reader for Writers* (pp. 401-407). New York, Oxford University Press.

O'Brien, M., & Gut, U. (2010). Phonological and phonetic realisation of different types of focus in L2 speech. In *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), 205-215. Frankfurt: Peter Lang.

O'Connor, J. D., & Arnold, G. K. (1973). *Intonation of Colloquial English.* London: Longman.

O'Grady, G.(2010). *A grammar of spoken English discourse: the intonation of increments*. London: Continuum.

Oakhill, J., & Kyle, F. (2000). The relation between phonological awareness and working memory. *Journal of Experimental Child Psychology, 75,* 152-164.

Ochs, E. (1992). Indexing Gender. In A. Duranti, & C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 335-58). Cambridge: Cambridge University Press.

Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, *38*, 1752-1775.

Ogden, R., & Routarinne, S. (2005). The vommunicative functions of final rises in Finnish intonation. *Phonetica*, *62*, 160-175.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1-18.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1-16.

Ohala, J., & Ewan, W. (1973). Speed of pitch change. *Journal of the Acoustical Society of America*, *53*, 345.

Ohara, Y. (1999). Performing gender through voice pitch: A cross-cultural analysis of Japanese and American English. In U. Pasero, & F. Braun (Eds.), *Wahrnehmung und Herstellung von Geschlecht: Perceiving and Performing Gender* (pp. 105-116). Wiesbaden: VS Verlag für Sozialwissenschaften.

Öhman, S. (1967). Word and sentence intonation: A quantitative model. *STL-QPSR*, *8*, 2-3.

Ordin, M., & Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech Language and Hearing Research*, *60*(6), 1493-1506.

Orrico, R., Cataldo, V., Savy, R., & Barone, L. (2016). Transfer, Fossilization and Prosodic Drift in Foreign Language Learning. In *Atti del XII Convegno Nazionale AISV*, *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere*, January *27-29 2016*, 117-132. Università di Salerno. DOI: 10.17469/O2102AISV000007

Ortega, L. (2009). *Understanding second language acquisition*. New York: Routledge.

Ortega, L., & Byrnes, H. (Eds.). (2008). *The Longitudinal Study of Advanced L2 Capacities*. *Second Language Acquisition Research*. New York; London: Routledge

Ortega-Llebaria, M., & Colantoni, L. (2014). L2 English intonation: Relations between form-meaning associations, access to meaning and L1-transfer. *Studies in second language acquisition, 36*, 331-353.

Ortiz Lira, H. (1994). *Contrastive analysis of English and Spanish sentence accentuation*. PhD thesis. University of Manchester.

Oxford, R. L., & Ehrman, M. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13,* 188-205.

Oxford, R. L., Nyikos, M., & Ehrman, M. (1988). Vive la différence? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals, 21*(4), 321-329.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(3), 261-283.

Paikeday, T. M. (1985). *The Native Speaker is Dead!*. Toronto: Paikeday.

Palková, Z. (1994). *Fonetika a fonologie češtiny.* Praha: Karolinium.

Palková, Z., & Volín, J. (2003). The role of F0 contours in determining foot boundaries in Czech. In *Proceedings of the 15th International Congress of Phonetic Sciences*, *August 3-9 2003*, M.J. Solé, D. Recasens, & J. Romero (Eds.), 1783-1786. Barcelona: Universitat Autònoma de Barcelona.

Palmer, H. (1922). *English Intonation, With Systematic Exercises*. Cambridge: Heffer.

Papoušek, M., & Hwang, S.-F. C. (1991). Tone and intonation in Mandarin baby talk to presyllabic infants: comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*, 481-504.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*(4), 2382-2393

Pardo, J. S. (2010). Expressing oneself in conversational interaction. In E. Morsella (Ed.), *Expressing oneself/expressing one's self: Communication, cognition, language, and identity* (pp. 183-196). London: Psychology Press.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183-195.

Park, C.-H., & Sim, K.-B. (2003). Emotion recognition and acoustic analysis from speech signal. In *Neural Networks*: *Proceedings of the International Joint Conference*, *July 20-24 2003*, 2594-2598. Portland, OR, USA. 10.1109/IJCNN.2003.1223240

Park, M. W. (2011). *Teaching Intonation Patterns through Reading Aloud*. MA (TESOL) thesis. Portland State University.

Pascoe, G. (1996). Pronunciation Analysis and Teaching. *PHONUS*, 2, 109-122. Saarbrücken: Institute of Phonetics, University of the Saarland.

Paterson, K. (2011). *A Handbook of spoken grammar: strategies for speaking natural English*. Peaslake: Delta Publ.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*(2), 449-472.

Patterson, D. (2000). *A Linguistic Approach to Pitch Range Modelling*. PhD thesis. University of Edinburgh.

Perdue, C. ed. (1993). *Adult language acquisition: Cross-linguistic perspectives. Volume 2. The results.* Cambridge: Cambridge University Press.

Pérez-Paredes, P. (2010). The death of the adverb revisited: Attested uses of adverbs in native and non-native comparable corpora of spoken English. In M. Moreno Jaén, F. Serrano Valverde, & M. Calzada Pérez (Eds.), *Exploring New Paths in Language Pedagogy. Lexis and Corpus-based Language Teaching* (pp. 157-172). London: Equinox.

Pérez-Paredes, P., & Bueno-Alastuey, M. C. (2019). A corpus-driven analysis of certainty stance adverbs: Obviously, really and actually in spoken native and learner English. *Journal of Pragmatics*, *140*, 22-32.

Pešková, A. (2017). Czech ToBI. *Phonetics and Phonology in Europe*, *June 12-15 2017*. Cologne, Germany. 10.13140/RG.2.2.36577.92007

Pešková, A., Colantoni, L., & Meisenburg, T. (2018). Initial and final intonation cues of Czech yes-no questions. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 50-54. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/27.html>. (18 June 2018).

Peters, J., Auer, P., Gilles, P., & Selting, M. (2015). Untersuchungen zur Struktur und Funktion regionalspezifischer Intonationsverläufe im Deutschen. Rückblick auf ein Forschungsprojekt. In R. Kehrein, A. Lameli, & S. Rabanus (Eds.), *Regionale Variation des Deutschen. Projekte und Perspektiven* (pp. 53-80). De Gruyter.

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, *57*(1), 108-146.

Pica, T. (1987). Second-Language Acquisition, social interaction, and the classroom. *Applied Linguistics*, *8*(1), 3-21.

Pica, T. (1994). Research on negotiation: What does it reveal about second language learning conditions, processes, outcomes? *Language Learning*, 44(3), 493-527.

Pica, T., Halliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, *11*, 63-90.

Pica, T., Holliday, L., Lewis, N., Berducci, D., & Newman, J. (1991). Language Learning Through interaction: What Role Does Gender Play? *Studies in Second Language Acquisition, 13,* 343-376.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, *35*, 233-255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and non-native speaker instructional discourse. *English for Specific Purposes*, *23*(1), 19-43.

Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics*, 6(2), 235-255.

Pickering, L. (2010). Intonation. In K. Malmkjaer (Ed.), *The Linguistics Encyclopedia*, 3rd edition (pp. 283-289). Routledge.

Pickering, L., & Levis, J. (2002). Assessing the Intonation Patterns of Second Language Learners. In *36th Annual TESOL Convention*, *April 9-13 2002*, Salt Lake City, Utah.

Pickering, L., & Litzenberg, J. (2011). Intonation as a Pragmatic Resource in ELF Interaction, Revisited. In A. Archibald, A. Cogo, & J. Jenkins (Eds.), *Latest Trends in ELF* (pp. 77-92). Cambridge Scholars.

Pickering, L., Hu, G., & Baker, A. (2012). The pragmatic function of intonation: Cueing agreement and disagreement in spoken English discourse and implications for ELT. In J. Romero-Trillo (Ed.), *Pragmatics*, *prosody & ELT* (pp. 199-218). Springer-Verlag.

Pickering, M. J., & Garrod, S. (2004). Toward a Mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences*, *27*, 169-226.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*, 203-228.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis. Cambridge, Mass.: MIT.

Pierrehumbert, J. (1981). Synthesizing intonation. *Journal of the Acoustical Society of America*, *70*(4), 985-995.

Pierrehumbert, J. B. (2003). Phonetic Diversity, Statistical Learning and Acquisition of Phonology. *Language and Speech*, *46*(2-3), 115-154.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge: MIT Press.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, MI: Univeristiy of Michigan Press.

Piller, I. (2001). Who, if anyone, is a native speaker?, *Anglistik*, *12*(2), 109-121.

Pincus, N. (2014). Teaching intonation with explicit instruction and audio-visual techniques. In *Proceedings 7th International Technology, Education and Development Conference (INTED), March 4-5 2013,* 3428-3435. Valencia, Spain.

Pirt, G. (1990). Discourse intonation problems for non-native speakers. In M. Hewings (Ed.), *Papers in discourse intonation* (pp. 145-156). Birmingham, AL: English Language Research.

Piske, T. (2012). Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 41-49). Springer.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*,191-215.

Pitzl, M.- L. (2005). Non-understanding in English as a Lingua Franca: Examples from a business context. *Vienna English Working Papers*, *14*(2), 50-71.

Podesva, R. (2011). Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics*, *39*(3), 233-264.

Polat, N. (2011). Gender differences in motivation and L2 accent attainment: An investigation of young Kurdish learners of Turkish. *Language Learning Journal*, *39*(1), 19-41.

Pon-Barry, H. (2008). Prosodic Manifestations of Confidence and Uncertainty in Spoken Language. In *Proceedings of 9th European Conference on Speech Communication and Technology Eurospeech*, *September 22-26 2008*, 74-77. Brisbane, Australia. <https://www.isca-speech.org/archive/interspeech_2008/i08_0074.html>. (17 July 2018).

Pon-Barry, H., & Shieber, S. (2009). Identifying Uncertain Words within an Utterance via Prosodic Features. In *Proceedings of 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *September 6-10 2009*, 1579-1582. Brighton, United Kingdom. <https://www.isca-speech.org/archive/interspeech_2009/i09_1579.html>. (02 November 2018).

Pon-Barry, H., Schultz, K., Bratt, E. O., Clark, B., & Peters, S. (2006): Responding to student uncertainty in spoken tutorial dialogue systems. *International Journal of Artificial Intelligence in Education*, *16*(2), 171-194.

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, *26*, 81-114.

Prechtel, C., & Clopper, C. G. (2016). Uptalk in Midwestern American English. In *Proceedings of Speech Prosody May 31-June 3 2016*, J. Barnes, & A. Brugos (Eds.), 133-137. Boston <https://www.isca-speech.org/archive/SpeechProsody_2016/pdfs/302.pdf >. (03 October 2018).

Preston, D. R. (1998). They speak really bad English down South and in New York City. In L. Bauer, & P. Trudgill (Eds.), *Language Myths* (pp. 139-149). Penguin: London.

Prieto, P., & Roseano, P. (coords). 2009-2013. Atlas interactivo de la entonación del español. http://prosodia.upf.edu/atlasentonacion/.

Puga, K. (2019). English intonation of advanced learners: A contrastive interlanguage analysis. In S. Götz, & J. Mukherjee (Eds.), *Learner Corpora and Language Teaching: Studies in Corpus Linguistics 92*, (pp. 191-217). John Benjamins.

Puga, K., Fuchs, R., Hudson, T., Setter, J., & Mok, P. (2018). The perception-production link in intonation: Evidence from German learners of English. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 685-689. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/177.html>. (17 June 2018).

Puga, K., Fuchs, R., Setter, J., & Mok, P. (2017). The Perception of English Intonation Patterns by German L2 speakers of English. In *Proceedings of 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *August 20-14 2017*, 3241-3245. Stockholm, Sweden. <https://www.isca-speech.org/archive/Interspeech_2017/abstracts/1279.html>. (15 August 2017).

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: a reexamination. L*anguage* L*earning*, *30*, 271-287.

Quilis, A. (1993). Tratado de Fonología y Fonéticas Espan˜olas. 2nd edition. Madrid: Gredos.

R Development Core Team (2015). *R: A language and environment for statistical computing*. RFoundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raith, J. (1986). Prosodie: Englisch als Fremdsprache, Deutsch als Basissprache. *Die neueren Sprachen*, *85*, 478-496.

Ramírez Verdugo, M. D. (2002). Non-native interlanguage intonation systems: a study based on a computerized corpus of Spanish learners of English. *ICAME Journal*; *26*, 115-132.

Ramírez Verdugo, M. D. (2003). *A Contrastive Analysis of Non-native Interlanguage English Intonation Systems and their Implication in the Organization of Information from a Functional Perspective. A Study based on a Computerized Corpus of Spanish Learners of English*. PhD thesis. Universidad Autónoma de Madrid, Madrid.

Ramírez Verdugo, M. D. (2005). The nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *Journal of Pragmatics, 37*(12), 2086-2115.

Ramírez Verdugo, M. D. (2006a). Prosodic realization of focus in the discourse of Spanish learners and English native speakers. *Estudios Ingleses de la Universidad Complutense*, *14*, 9-32.

Ramírez Verdugo, M. D. (2006b). Non-native intonation: What information does it transmit? In A. María Hornero, M. J. Luzón, & S. Murillo (Eds.), *Corpus Linguistics. Applications for the Study of English, 25*, 507-526.

Ramírez Verdugo, M. D. (2008). A cross-linguistic study on the pragmatics of intonation in directives. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 205-234). Berlin: Mouton de Gruyter.

Ramírez Verdugo, M. D., & Romero-Trillo, J. (2005). The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics*, *2*(2), 151-168.

Rampton, B. (2009). Interaction ritual and not just artful performance in crossing and stylization, *Language in Society*, *38,* 149-176.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahier s de linguistique française*, *28*, 41-66.

Ray, G. B., & Zahn, C. J. (1990). Regional speech rates in the United States: A preliminary analysis. *Communication Research Reports*, *7*, 34-37.

Reichel, U. D. (2015). Personality Prediction Based on Intonation Stylization. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, *August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.), 616-620. Glasgow, Scotland, UK: The University of Glasgow.

<https://www. internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0616. pdf>. (05 August 2018).

Reitter, D., & Moore, J. D. (2007). Predicting success in dialogue. *Annual Meeting - Association for Computational Linguistics*, *45*, 808-815.

Reyelt, M., Grice, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Prosodische Etikettierung des Deutschen mit ToBI. In D. Gibbon (Ed.), *Natural Language and Speech Technology, Results of the third KONVENS conference* (pp. 144-155). Berlin, New York: Mouton de Gruyter.

Riesco-Bernier, S. (2012). Same but different: the pragmatic potential of native vs. non-native teachers' intonation in the EFL classroom. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 171-198). Dordrecht: Springer.

Riesco-Bernier, S., & Romero-Trillo, J. (2008). The acoustics of 'newness' and its pragmatic implications in classroom discourse. *Journal of Pragmatics*, *40*, 1103-1116.

Ritchart, A., & Arvaniti, A. (2014). The form and use of uptalk in Southern California English. In *Proceedings to the 7th Speech Prosody Conference, May 20-23 2014*, N. Campbell, D. Gibbon, & D. Hirst (Eds.), 331-335. Trinity College Dublin, Ireland. (26 August 2017). <http://www.sprosig.org/sp2014/sp7book.pdf>. (05 July 2015).

Roach, P. (2009 [1991]). *English Phonetics and Phonology: A practical course*. Cambridge: Cambridge University Press.

Robb, M. P., Maclagan, M. A., & Chen, Y. (2004). Speaking rates of American and New Zealand varieties of English. *Clinical Linguistics and Phonetics*, *18*, 1-15.

Rodgers, J. (2000). The phonatory correlates of juncture in German. In *Proceedings of the 5th Seminar on Speech Production: Models and Data*, *May 1-4 2000*, 289-292. Kloster Seeon, Baravia.

Rogerson-Revell, P. (2011). *English Phonology and Pronunciation Teaching*. Bloomsbury.

Romero-Trillo, J. (2001). A mathematical model for the analysis of variation in discourse. *Journal of Linguistics*, *37*(3), 527-550.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, *34*, 769-784.

Romero-Trillo, J. (2007). Adaptive management in discourse: The case of involvement discourse markers in English and Spanish conversations. *Catalan Journal of Linguistics*, *6*, 81-94.

Romero-Trillo, J. (Ed.). (2012). *Pragmatics and Prosody in English Language Teaching*. Dordrecht: Springer.

Romero-Trillo, J. (2014) 'Pragmatic punting' and prosody: Evidence from corpora. In M. A. Gómez González, F. J. Ruíz de Mendoza Ibáñez, F. Gonzálvez García, & A. Downing (Eds.), *The Functional Perspective on Language and Discourse: Applications and Implications* (pp. 209-221). Amsterdam: John Benjamins.

Romero-Trillo, J. (2018). Corpus pragmatics and second language pragmatics: a mutualistic entente in theory and practice. *Corpus Pragmatics*, *2*(2), 113-127.

Romero-Trillo, J. (2019). Prosodic pragmatics and feedback in intercultural communication. *Journal of Pragmatics, 151*, 91-102.

Romero-Trillo, J., & Lenn, E. (2011). Do you "(mis)understand" what I mean? Pragmatic Strategies to Avoid Cognitive Maladjustment. *Journal of English Studies*, *9*, 223-241.

Romero-Trillo, J., & Llinares-García, A. (2004). Prosodic competence in reading aloud: an acoustic corpus-based study of native and non-native (Spanish) speakers of English. *Estudios Ingleses de la Universidad Complutense*, *12*, 63-77.

Romero-Trillo, J., & Maguire, L. (2011). Adaptive context, the fourth element of meaning. *International Review of Pragmatics*, *3*, 226-239.

Romero-Trillo, J., & Newell, J. (2012). Prosody and feedback in native and non-native speakers of English. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 117-132). Dordrecht: Springer.

Rosen, A. (2019). The use of smallwords in the speech of German learners of English. In S. Götz, & J. Mukherjee (Eds.), *Learner Corpora and Language Teaching* (pp. 220-241). Amsterdam: John Benjamins.

Rüdiger, S. (2016). Cuppa coffee? Challenges and opportunities of compiling a conversational English corpus in an Expanding Circle setting. In H. Christ, D. Klenovšak, L. Sönning, & V. Werner (Eds.), *A Blend of MaLT: Selected Contributions from the Methods and Linguistic Theories Symposium 2015* (pp. 49-71). Bamberg: University of Bamberg Press.

Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 15-21). New York: Irvington Publishers.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696-735.

Sagisaka, Y., Campbell, N., & Higuchi, N. (Eds.). (1997). *Computing Prosody: Computational Models for Processing Spontaneous Speech*. New York: Springer.

Saito, K. (2013). Age effects on late bilingualism: The production development of /r/ by highprofciency Japanese learners of English. *Journal of Memory and Language*, *69*(4), 546-562.

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563-595.

Saito, K. (2018). Advanced Second Language Segmental and Suprasegmental Acquisition. In P. A. Malovrh, & A. G. Benati (Eds.), *The Handbook of Advanced Proficiency in Second Language Acquistion* (pp. 282-303). Wiley.

Saito, Y., & Saito, K. (2016). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, *21*(5), 1-20.

Sánchez-Alvarado, C. (2020). Syntactic and Prosodic Marking of Subject Focus in American English and Peninsular Spanish. In A. Morales-Front, M. J. Ferreira, R. P. Leow, and C. Sanz (Eds.), *Hispanic Linguistics: Current issues and new directions* (pp. 184-203). John Benjamins.

Saville-Troike, M., & Barto, K. (2017). *Introduction Second Language Acquisition*. 3rd edition. Cambridge: Cambridge University Press.

Scarborough, R. A., Keating, P., Mattys, S. L., T Cho, & A. Alwan (2009). Optical phonetics and visual perception of lexical and phrasal stress in English. *Language and Speech*, 52(2-3), 135-175.

Scharff-Rethfeldt, W., Miller, N., & Mennen, I. (2008). Unterschiede in der mittleren Sprechtonhöhe bei Deutsch/Englisch bilingualen Sprechern. *Sprache, Stimme, Gehör 32*, *32*(3), 123-128.

Schegloff, E. A. (1981). Discourse as an interactionl achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Georgetown University Round Table on Languages and Linguistics Washington D.C.* (pp. 71-93). Georgetown University Press.

Schegloff, E. A. (1996). Turn-organization: one intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar*, (pp. 52-133). Cambridge: Cambridge University Press.

Schilling, N. (2013). *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press. Print.

Schütze, U. (2016). *Language Learning and the Brain. Lexical Processing in Second Language Acquisition*. Cambridge: Cambridge University Press.

Schweitzer, A., & Lewandowski, N. (2013). Convergence of Articulation Rate in Sponatenous Speech. In *Proceedings of 14th Annual Conference of the International Communication Association (INTERSPEECH), August 25-29 2013*, 525-529. Lyon, France. <https://www.isca-speech.org/archive/interspeech_2013/i13_0525.html>. (01 February 2017).

Schweitzer, K., Walsh, M., & Schweitzer, A. (2017). To see or not to see: Interlocutor visibility and likeability influence convergence in intonation. In *Proceedings of 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *August 20-14 2017*, 919-923. Stockholm, Sweden. <https://www.isca-speech.org/archive/Interspeech_2017/abstracts/1248.html>. (26 August 2017).

Scovel, T. (1988). *A Time to Speak. A Psycholinguistic Inquiry to the Critical Period for Human Speech*. Rowley MA: Newbury House.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213-223.

Scuffil, M. (1982). *Experiments in Comparative Intonation*. Tübingen. Niemeyer.

Seliger, H.W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W.C. Ritchie (Ed.), *SLA Research: Issues and Implications* (pp. 11-19). New York: Academic Press.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(3), 209- 231.

Selinker, L. (1992). *Rediscovering Interlanguage*. London: Longman.

Selinker, L., & Lamendella, J. (1978). Two perspectives on fossilization in interlanguage learning. *Interlanguage Studies Bulletin*, *3*(2), 143-191.

Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, *29*, 477-517.

Selting, M. (2005). Syntax and prosody as methods for the construction and identification of turnconstructional units in conversation. In A. Hakulinen, & M. Selting (Eds.), *Syntax and Lexis in conversation. Studies on the use of linguistic resources in talk-in-interaction* (pp. 17-44). Amsterdam, Philadelphia: Benjamins.

Selting, M. (2010). Prosody in interaction: State of the art. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), *Prosody in Interaction* (pp. 3-40). Amsterdam: John Benjamins.

Setter, J., Wong, C. S. P., & Chan, B. H. S. (2010). *Hong Kong English.* Edinburgh: Edinburgh University Press.

Sheldon A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *19*, 243-261.

Shevchenko, T. I., & Skopintseva, T. S. (2004). Prosody Variation in English: Geographical, Social, Situational. In *Proceedings of Speech Prosody*, *March 23-26, 2004, Nara, Japan,* B. Bel, & I. Marlien (Eds.), 139-142. <https://www.isca-speech.org/archive/sp2004/papers/sp04_139.pdf>. (05 July 2018).

Shobbrook, K., & House, J. (2003). High rising tones in Southern British English. In *Proceedings of the 15th international Congress of Phonetic Sciences*, *August 3-9 2003*, M. J. Sole, D. Recansens, & J. Romero (Eds.), 1273-1276. Barcelona: Casual Productions.

Shokeir, V. (2008). Evidence for the stable use of uptalk in South Ontario English. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 16-24.

Shriberg, L. D., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Resarch*, *27*, 456-65.

Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptuallearning of systematic variation in Spanish-accented speech, *Journal of the Acoustical Society of America*, *125*(5), 3306-3316.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing*, *October 13-16 1992*, 867-870. Banff, Alberta, Canada. <https://www.isca-speech.org/archive/archive_papers/icslp_1992/i92_0867.pdf>. (2 Januar 2015).

Šimáčková, Š., Kolářová, K., & Podlipský, V. J. (2014). Tempo and connectedness of Czech-accented English speech. *Concordia Working Papers in Applied Linguistics, 5*, 667-677.

Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *Acta Universitatis Carolinae. Philologica 2/ Phonetica Pragensia*, 109-128.

Skarnitzl, R., & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of common Czech. *Acta Universitatis Carolinae Philologica 3 /Phonetica Pragensia*, 7-17.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Langauge Teaching Research*, *1*, 185-211.

Smith, C. L. (2007). Prosodic accommodation by French speakers to a nonnative interlocutor. In *Proceedings of the XVIth International Congress of Phonetic Sciences, August 6-10 2007,* W. J. Barry, & J. Trouvain (Eds.), 1081-1084. Universität des Saarlandes, Saarbrücken, Germany.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25-38.

Smorenburg, L., Rodd, J., & Chen, A. (2015). The effect of explicit training on the prosodic production of L2 sarcasm by Dutch learners of English. In *Proceedings of the 18th*

International Congress of Phonetic Sciences (ICPhS), August 10-14 2015, M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0959.pdf>. (05 September 2018).

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103-130.

Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, *58*(2), 574-608.

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. L*anguage & Speech*, *20*(4), 357-365.

Snow, D., & Balog, H. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, *112*, 1025-1058.

Sonsaat, S. (2017). The influence of an online pronunciation teacher's manual on teachers' cognitions. Unpublished PhD thesis, Iowa State University.

Sosa, J. M. (1999). *La entonación del español: su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.

Speer, S. R., Warren, P., & Schafer, A. J. (2003). Intonation and sentence processing. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, M. Solé, D. Recasens and J. Romero (Eds.), 95-105. Barcelona.

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35-98.

Speeter Beddor, P. (2015). The relationship between language users' perception and production repertoires. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS1041.pdf>. (04 September 2018).

Spezzini, S. (2004). English immersion in Paraguay: Individual and sociocultural dimensions of language learning and use. *International Journal of Bilingual Education and Bilingualism*, *7*, 412-431.

Spolsky, B. (1989). *Conditions for Second Language Learning: Introduction to a General Theory.* Oxford: Oxford University Press.

Staples, S. (2015). *The discourse of nurse-patient interactions: contrasting the communicative styles of U.S. and international nurses*. Philadelphia: John Benjamins.

Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genre, and language ratings. *Journal of Second Language Writing*, *32*, 17-35.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and academic writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225.

Staples, S., LaFlair, G. T., & Egbert, J. (2017). Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *The Modern Language Journal*, *101*(1), 1-20.

Steinberg, D. D, & Sciarini, N. (2006). *Introduction to Psycholinguistics*. 2nd edition. London: Longman.

Stirling, L., Fletcher, J., Mushin, I., & Wales, R. (2001). Representational issues in annotationg: using the Australian map task corpus to relate to prosody and discourse structure. *Speech Communication*, *3*, 113-134.

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and perceptual indicators of emotional stress. *Journal of the Acoustic Society of America*, *73*(4), 1354-1360.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. L*anguage* Le*arning*, *26*, 233-253.

Svartvik, J. (Ed.). (1990). The London Corpus of Spoken English: Description and Research. *Lund Studies in English*, *82*. Lund: Lund University Press.

Svartvik, J. (1991). What can real spoken data teach teachers of English? In J. E. Alatis (Ed.), *Georgetown University Round Table of Languages and Linguistics. Linguistics and language pedagogy: the state of the art*, (pp. 555-566).

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, *101*, 485-532.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language & Speech*, *37*(1), 21-43.

Swerts, M., & Hirschberg, J. (2010). Prosodic predictors of upcoming positive or negative content in spoken messages. *The Journal of the Acoustical Society of America*, *128*(3), 1337-1345.

Swerts, M., & Krahmer, E. (2004). Congruent and incongruent audiovisual cues to prominence. In *Proceedings of Speech Prosody*, *March 23-26, 2004, Nara, Japan,* B. Bel, & I. Marlien (Eds.), 69-72. <https://www.isca-speech.org/archive/sp2004/papers/sp04_069.pdf>. (05 July 2018).

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*, 81-94.

Szczepek Reed, B. (2010). Intonation phrases in natural conversation. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), *Prosody in Interaction* (pp. 191-212). Amsterdam: John Benjamins.

't Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics, 3*, 235-55.

Tahta, S., Wood, M., & Loewenthal, K. (1981a). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech, 24*(3), 265-272.

Tahta, S., Wood, M., & Loewenthal, K. M. (1981b). Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech*, *24*(4), 363-372.

Tannen, D. (1986). Cross-cultural Communication. In T. van Dijk (Ed.), *Handbook of Discourse Analysis, vol. 4, Discourse Analysis in Society* (pp. 203-215). London & Orlando: Academic Press.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning, 58*(2), 439-473.

Taylor, D. S. (1993). Intonation and accent in English: what teachers need to know. *International Review of Applied Linguistics in Language Teaching*, *31*(1), l-22.

Taylor, P. (2000). *Analysis and Synthesis of Intonation using the Tilt Model*. Technical report, Centre for Speech Technology Research, University of Edinburgh.

Tench, P. (1996). *The intonation systems of English*. London: Cassell.

Tench, P. (2003). Processes of semogenesis in English intonation. *Functions of Language*, *10*(2), 209-234.

Tench, P. (2015). *The intonation systems of English*. London: Bloomsbury Academic.

Terraschke, A., & Holmes, J. (2007). Und tralala: vagueness and general extenders in German and New Zealand English. In J. Cutting (Ed.), *Vague Language Explored* (pp. 198-220). London: Palgrave Macmillan.

Thomason, J., Nguyen, H. V., & Litman, D. (2013). Prosodic entrainment and tutoring dialogue success. *Artificial Intelligence in Education*, *7926*, 750-753.

Thompson, I. (1991). Foreign accents revisited: the English pronunciation of Russian immigrants. L*anguage Learning*, *41*(2), 177-204.

Toivanen, J. (2003). Tone choice in the English intonation of proficient non-native speakers. In *PHONUM 9: Proceedings from Fonetik 2003*, M. Heldner (Ed.), 1-4.

Tomlinson, J. M. Jr., & Fox Tree, J. E. (2009). Talking it up: How the function of rising declaratives depends on pronolongations and listeners' expectations. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, *July 29 - August 1 2009*, N. Taatgen (Ed.), 1234-1239. Amsterdam, Netherlands.

Tomlinson, J. M. Jr., & Fox Tree, J. E. (2011). Listeners' comprehension of uptalk in spontaneous speech. *Cognition*, *119*(1), 58-69.

Torres, L. (2002). Bilingual discourse markers in Puerto Rican Spanish. *Language in Society, 31*(1), 65- 83.

Tortel, A. (2013). Prosody in a contrastive learner corpus. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpora*. *Studies in Corpus Linguistics Series 59* (pp. 227-248). Amsterdam: John Benjamins.

Tortel, A., & Hirst, D. (2010). Rhythm metrics and the production of English L1/L2. In *Proceedings of Speech Prosody*, *May 11-14 2010*, M. Hasegawa-Johnson (Ed.), paper 959. Chicago, IL, USA. <https://www.isca-speech.org/archive/sp2010/papers/sp10_959.pdf>. (21 April 2015).

Tottie, G. (2002). *An Introduction to American English*. Blackwell Publishers.

Tottie, G., & Hoffmann, S. (2006). Tag questions in British and American English. *Journal of English Linguistics*, *34*(4), 283-311.

Towell, R. (2002). Relative degrees of fluency. A comparative case study of advanced learners of French. *International Review of Applied Linguistic in Language*, *40*, 117-150.

Tracy-Ventura, N., & Paquot, M. (Eds.). (2021). *Routledge Handbook of Second Language Acquisition and Corpora*. London/New York: Routledge.

Trafton, A. (2018). Cognitive scientists define critical period for learning language: Study shows children remain adept learners until the age of 17 or 18. MIT NEWS Office.

<http://news.mit.edu/2018/cognitive-scientists-define-critical-period-learning-language-0501>. (15 July 2018).

Trager, G., & Smith, H. (1951). *An Outline of English Structure*. Washington: American Council of Learned Societies.

Trim, J. (1964). Tonetic stress marks for German. In D. Abercrombie, D. Fry, P. MacCarthy, N. C. Scott, & J. Trim (Eds.), *In Honour of Daniel Jones* (pp. 374-383), London: Longman.

Trim, J. (1988). Some contrastive intonated features of British English and German. In J. Klegrafand, & D. Nehls (Eds.), *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel's 60th Birthday* (pp. 235-249), Heidelberg. Julius Groos.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect if L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Trofimovich, P., & Baker, W. (2007). Learning prosody and fluency characteristics of L2 speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, *28*, 251-276.

Trouvain, J. (2001). Phonetic aspects of "speech-laughs". In *Proceedings of the Conference on Orality & Gestuality (ORAGE 2001)*, *June 18-22 2001*, K. Miligan (Ed.), 634-639. Aix-en-Provence, France.

Trouvain, J., & Grice, M. (1999). The effect of tempo on prosodic structure. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99)*, *Vol. 1, August 1-7 1999,* J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), 1067-1070. San Francisco, United States.

Trouvain, J., Zimmerer, F., Möbius, B., Gósy, M., & Bonneau, A. (Eds.). (2017). *Segmental, prosodic and fluency features in phonetic learner corpora: special issue*. Amsterdam: John Benjamins.

Truckenbrodt, H. (2007). Upstep on edge tones and on nuclear accents. In C. Gussenhoven, & T. Riad (Eds.), *Tones and Tunes, Volume 2: Experimental Studies in Word and Sentence Prosody* (pp. 349-386). Berlin: De Gruyter Mouton.

Tyler, J. (2012). *Discourse Prosody in Production and Perception*. PhD thesis, University of Michigan.

Tyler, J. (2014). Rising Pitch and Quoted Speech in Everyday American English. In *Proceedings to the 7th Speech Prosody Conference, May 20-23 2014*, N. Campbell, D. Gibbon, & D. Hirst (Eds.), 548-552. Trinity College Dublin, Ireland. (27 August 2017). <https://www.isca-speech.org/archive/SpeechProsody_2014/pdfs/101.pdf>. (05 July 2015).

Tyler, J. (2015). Expanding and mapping the indexical field: Rising pitch, the uptalk stereotype, and perceptual variation. *Journal of English Linguistics*, *43*(4), 284-310.

Tyler, J. (2019, *under revision*) *Gender, questions and phonetic variation in uptalk: A corpus analysis of terminal rising pitch in the Santa Barbara Corpus of Spoken American English.* Manuscript.

Ulbrich, C. (2004). A comparative study of declarative utterances in Swiss and German standards. In P. Gilles, & J. Peters (Eds.), *Regional Variation in Intonation* (pp. 99-122). Max Niemeyer Verlag.

Ulbrich, C. (2006a). Pitch Range Is Not Pitch Range. In *Proceedings of Speech Prosody, May 2-5 2006,* R. Hoffmann, & H. Mixdorff, paper 041. Dresden, Germany. <https://www.isca-speech.org/archive/sp2006/papers/sp06_041.pdf>. (04 April 2015).

Ulbrich, C. (2006b). Prosodic Phrasing in Three German Standard Varieties. *University of Pennsylvania Working Papers in Linguistics*, *12*(1), 361-373.

Ulbrich, C. (2008). Acquisition of regional pitch patterns in L2. In *Proceedings of the 4th Speech Prosody, May 6-9 2008,* P. A. Barbosa, S. Madureira, & C. Reis (Eds.), 575-578. Campinas, Brazil. <https://www.isca-speech.org/archive/sp2008/papers/sp08_575.pdf>. (15 May 2016).

Ulbrich, C. (2010). Belfast Intonation in L2 speech. In *Proceedings of Speech Prosody*, *May 11-14 2010*, M. Hasegawa-Johnson (Ed.), paper 963. Chicago, IL, USA. <https://www.isca-speech.org/archive/sp2010/papers/sp10_963.pdf>. (21 April 2015).

Ulbrich, C. (2012). German pitches in English: Production and perception of cross-varietal differences in L2. Computational Modeling of Bilingualism, 16(2), 397-419.

Ulbrich, C. (2015). Developmental Stages and Variability in the Acquisition of Second Language Segments and Prosody. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0549.pdf>. (03 August 2018).

Ulbrich, C., & Mennen, I. (2015). When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism*, *20*(5), 522-549.

Uličný, O., & Prošek, M. (Eds.). (2013). *Studie k Moderní mluvnicičeštiny - K české fonetice a pravopisu.* UPOL: University of Olomouc.

Ullakonoja, R. (2007). Comparison of pitch range in Finnish (L1) and Russian (L2). In *16th International Congress of Phonetic Sciences (ICPhS), August 6-10 2007*, 1701-1704. Saarbrücken, Germany.

Ullman, M. T., Miranda, R. A., & Travers, M. L. (2007). Sex differences in the neurocognition of language. In J. B. Becker, K. J. Berkley, N. Geary, E. Hampson, J. P. Herman, & E. Young (Eds.), *Sex Differences in the Brain: From Genes to Behavior* (pp. 291-309). Oxford Scholarship Online.

Urbani, M. (2012). Pitch range in L1/L2 English. An analysis of f0 using LTD and linguistic measures. *In Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*, M. G. Busà, & A. Stella (Eds.), 79-83. University of Padova, Italy.

Uther, M., Knoll, M., & Burnham, D. (2007). Do you speak E-N-G-L-I-S-H? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*, 2-7.

Uzal, M., Peltonen, T., Huotilainen, M., & Aaltonen, O. (2015). Degree of perceived accent in Finnish as a second language for Turkish children born in Finland. *Language Learning, 65*(3)*,* 477-503.

Van Bezooijen, R. (1995). Sociocultural aspects of Pitch differences between Japanese and Dutch women. *Lang Speech*, *38*(3), 253-265

Van Els, T., & de Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal*, *71*, 147-155.

Varonis, E. M., & Gass, S. (1985). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, *6*(1). 71-90.

Vogel, I. (1991). Prosodic phonology: second language acquisition data as evidence in theoretical phonology. In T. Huebner, & C.A. Ferguson (Eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories* (pp. 47-65). Amsterdam: John Benjamins.

Volín, J. (2008). Z intonace čtených zpravodajství: výška první slabiky v taktu. Čeština doma a ve světě, 2008 (1-2), 89-96.

Volín, J., & Galeone, D. (2017). Prosodic phrases in native and foreign-accented speech. Accents 2017. Accents in use. Łódź, Poland.

Volín, J., & Galeone, D. (2018). Prosodic breaks in foreign-accented speech. Accents 2018. Accents go to School. Łódź, Poland.

Volín, J., & Johaníková, T. (2018). Weak structural words in British and Czech English. In: Volín, J. & Skarnitzl, R. (Eds.), *The Pronunciation of English by Speakers of Other Languages*, 181-195. Newcastle upon Tyne: Cambridge Scholars Publishing.

Volín, J., & Skarnitzl, R. (2007). Temporal downtrends in Czech read speech. In *Proceedings of 8th European Conference on Speech Communication and Technology (INTERSPEECH)*, *August 27-31 2007*, 442-445. Antwerp, Belgium. <https://www.isca-speech.org/archive/interspeech_2007/i07_0442.html>. (02 April 2017).

Volín, J., & Skarnitzl, R. (2010a). Suprasegmental Acoustic Cues of Foreigness in Czech English. In *Proceedings of Odyssey 2010: The Speaker and Language Recognition Workshop*, *June 28 - July 1 2010*, 271-278. Brno, Czech Republic. <https://www.isca-speech.org/archive_open/archive_papers/odyssey_2010/papers/od10_045.pdf>. (07 June 2015).

Volín, J., & Skarnitzl, R. (2010b). The strength of foreign accent under adverse listening conditions. *Speech Communication*, *52*(11-12), 1010-1021.

Volín, J., Galeone, D., & Johnson, W. (2017). Pitch range of intonation contours in English Czech. *Philologica* 3. 55-64.

Volín, J., Poesová, K., & Weingartová, L. (2015). Speech melody properties in English, Czech and Czech English: Reference and interference. *Research in Language*, 13(1), 107-123.

Volín, J., Skarnitzl, R., & Henderson, A. (2018). Perceptual Impact of Foreign-Accented Speech. In eds. Jan Volín & Radek Skarnitzl. The Pronunication of English by Speakers of Other Languages. Cambridge Scholars Publishing. 73-94.

Volín, J., Weingartová, L., & Niebuhr, O. (2014). Between Recognition and Resignation - The Prosodic Forms and Communicative Functions of the Czech Confirmation Tag "jasně". Speech Prosody 7, 115-119.

Volín, J., Weingartová, L., & Niebuhr, O. (2016). The Prosody of the Czech Discourse Marker '*Jasně*': An Analysis of Forms and Functions". Phonetica 73(3-4): 314-337.

Vykouková, M. (2014). *The Approach of Czech Upper-secondary English Teachers to Pronunciation Teaching*. (Unpublished Master's Thesis). Faculty of Arts, Charles University, Prague.

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task, *Frontiers in Psychology* 6, 1-18.

Walker, R. (1999). Proclaimed and perceived wants and needs among Spanish teachers of English. Speak Out! 24.1: 25-32.

Walsh, T., & Diller, K. (1981). Neurolinguistic considerations on the optimal age of second language learning. in K. Diller (ed.): Individual Differences and Universals in Language Learning Aptitude. Rowley, Mass.: Newbury House.

Ward, G., & Hirschberg, J. (1985). Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation. In: *Language* 61 (4), 747-776.

Ward, N. G., & Gallardo, P. (2015). A Corpus for Investigating English-Language Learners' Dialog Behaviors, UTEP Technical Report. 1-8.

Ward, N. G., & Gallardo, P. (2017). Non-native differences in prosodic-contstruction use. Dialogue & Discourse 8 (1), 1-30.

Ward, N. G. (2015). Midlevel prosodic features toolkit. https://github.com/nigelgward/midlevel.

Warren, P. (2005). Patterns of Late Rising in New Zealand English. *Language Variation and Change* 17 (2). 209-30.

Warren, P. (2016). *Uptalk.The phenomenon of rising intonation*. Cambridge: Cambridge University Press.

Warren, P., & Britain, D. (2000). Intonation and Prosody in New Zealand English. In Allan Bell & Koenraad Kuiper. New Zealand English. Amsterdam [u.a.]: John Benjamins. 146-172.

Warren, P., & Daly, N. (2005). Characterizing New Zealand English intonation: Broad and narrow analysis. In Allan Bell, R. Harlow & D. Sparks (Eds.), *Languages of New Zealand*. Wellington: Victoria University Press. 217-237.

Weingartová, L., Churaňová, E., & Šturm, P. (2014b). Transitions, pauses and overlaps: Temporal characteristics of turn-taking in Czech. Speech Prosody 7, 502-506.

Weingartová, L., Poesová, K., & Volín, J. (2014a). Prominence Contrasts in Czech English as a Predictor of Learner's Proficiency. In Campbell, Gibbon & Hirst (Eds.) Speech Prosody 7, 2014. 236-240.

Weirich, M., & Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. PLoS ONE 13(12): p. 1-18.

Wells, J. C. (2006). *English intonation: an introduction.* Cambridge: Cambridge University Press.

Wells, J. C. (2014). *Sounds Interesting: observations on English and general phonetics*. Cambridge: Cambridge University Press.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. Applied Linguistics; 15(4):399-420.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse, *Studies of Second Language Acquisition* 20, 1-25.

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press.

Wichmann, A. (2000). Intonation in Text And Discourse: Beginnings, Middles,and Ends. Harlow, England: Longman.

Wichmann, A., & Caspers, J. (2001). Melodic cues to turntaking in English: evidence from perception. In J. van Kuppeveldt & R. Smith (Eds.), *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, 180-5.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag: New York.

Widdowson, H. G. (1994). The ownership of English, *TESOL Quarterly* 28 (2), 377-389.

Wilhelm, S. (2016). Towards a typological classification and description of HRTs in a multidialectal corpus of contemporary English. Speech Prosody 2016 Proceedings. 138-142.

Willems, N. (1982). *English Intonation from a Dutch Point of View*. Ambach: Intercontinental Graphics.

Williams, B. (1996a). The status of corpora as linguistic data. In: G. Knowles & P.R. Alderson (Eds.), *Working with Speech: The Computational Analysis of Formal British EnglishSpeech.* London: Longmans.

Williams, B. (1996b). The formulation of an intonation transcription system for British English. In: G. Knowles & P.R. Alderson (Eds.), Working with Speech: The ComputationalAnalysis of Formal British English Speech. London: Longmans.

Williams, M. J., & Tiedens, L. Z. (2016). The Subtle Suspension of Backlash: A Meta-Analysis of Penalties for Women's Implicit and Explicit Dominance Behavior. Psychological Bulletin 142 (2). 165-197.

Willis, E.W. (2010). Dominican Spanish Intonation. In P. Prieto & P. Roseano (Eds.), Transcription of Intonation of the Spanish Language. Munich: Lincom Europa, 123-153.

Wolfson, N. (1976). Speech events and natural speech: Some implications for sociolinguistic methodology. *Language in Society* 5(2): 189-209.

Wolfson, N. (1986a). Research methodology and the question of validity. *TESOL Quarterly,* 20 (4), 689-699.

Wolfson, N. (1986b). The bulge: A theory of speech behavior and social distance. Working papers in educational linguistics 2 (1): 55-83.

Wolfson, N. (1989). *Perspectives: Sociolinguistics and TESOL.* Cambridge, MA: Newbury House.

Wollermann, C. (2012). Prosodie, nonverbal Signale, Unsicherheit und Kontext - Studien zur pragmatischen Fokusinterpretation. Dissertation Universität Duisburg-Essen.

Yan, X., & Staples, S. (2019). Fitting MD analysis in an argument-based validity framework for writing assessment: Explanation and generalization inferences for the ECPE. Language Testing. 1-26. Press.

Yang, L.-C. (2004). Duration and pauses as cues to discourse boundaries in speech, *Proceedings of Speech Prosody* 2004, Nara, Japan. 267-270.

Yngve, V. H. (1970). On Getting a Word in Edgewise, Papers from the Sixth Regional Meeting Chicago Linguistic Society (Chicago: Chicago Linguistic Society), 567-577.

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits, PloS One 8(9), e74746.

Yu, G. (2009). Lexical Diversity in Writing and Speaking Task Performances, Applied Linguistics 31/2: 236-259.

Yuan, J., Shen, L., & Chen, F. (2002). The acoustic realization of anger, fear, joy and sadness in Chinese, in Proceedings ICSLP-2002, Denver, CO, 2025-2028.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, 52, 33-47.

Zárate-Sández, G. (2015). Percepetion and production of intonation among English-Spanish bilingual speakers at different proficiency levels, Georgetown University, 2015.

Zárate-Sández, G. (2016). Categorical perception and prenuclear pitch peak alignment in Spanish, in *Proceedings of Speech Prosody 2016*, pp. 663-667.

Zárate-Sández, G. (2017). Reexamining Foreign Accent: How Much Can Personality Explain?. *Ilha do Desterro*, 70(3), 227-243.

Zárate-Sández, G. (2018). Production of final boundary tones in declarative utterances by English speaking learners of Spanish. In *Proceedings of the 9th Internationational Conference on Speech Prosody*, *June 13-16 2018,* K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), 927-931. Poznań, Poland. <https://www.isca-speech.org/archive/SpeechProsody_2018/abstracts/192.html>. (18 June 2018).

Zellers, M., & Schweitzer, A. (2017). An investigation of pitch matching across adjacent turns in a corpus of spontaneous German. In *Proceedings of 18th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, *August 20-14 2017*, 2336-2340. Stockholm, Sweden. <https://www.isca-speech.org/archive/Interspeech_2017/abstracts/0811.html>. (16 September 2017).

Zellner, B. (1994). Pauses and the temporal structure of speech, in E. Keller [Ed.], Fundamentals of speech synthesis and speech recognition, 41-62, Chichester: John Wiley.

Zimmerer, F., Andreeva, B., Jügler, J., & Möbius, B. (2015). Comparison of pitch profiles of German and French speakers speaking French and German. In *18th International Congress of Phonetic Sciences (ICPhS), August 10-14 2015,* M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith, & J. Scobbie (Eds.). Glasgow, Scotland, UK: The University of Glasgow. <https://www. international phoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0183.pdf>. (02 September 2018).

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. Proc. Speech Prosody 7, Dublin, Ireland, 1037-1041. A. Cruttenden, *Gimson's Pronunciation of English (8th edition)*. London: Routledge.

Zubizarreta, M. L., & Nava, E. (2011). Encoding discourse-based meaning: Prosody vs. syntax. Implications for second language acquisition. *Lingua*, *121*, 652-669.

Zuengler, J. (1991). Accommodating in Native-Non-native Interactions: Going Beyond the "What" to the "Why" in Second-Language Research. In Giles, Howard; Coupland, Justine; Coupland, N. Contexts of Accommodation. New York: Cambridge University Press.

Zvonik, E., & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration, Proceedings of Eurospeech 2003, Geneva, Switzerland: 777-780.

Appendices

Appendix 1. Detailed manual for the prosodic annotation of spontaneous speech

The prosodic transcription system used for this project is a slightly modified version of the ToBI annotation system. The ToBI model has been chosen as explained in Section 3.4 because it is the most widely used annotation system, which has been successfully applied to both GA English and AusE (cf. Warren 2016: 70). The instructions in this manual were used to train annotate the corpus samples (see Section 3.3 for the selected corpus samples).

Each recording is annotated according to the following 10 tiers:

Task	Interview task tier	IT
SA	Speech acts tier	IT
IP	Intonation phrase and orthographic tier	IT
ip	Intermediate phrase and orthographic tier	IT
HL	Tone tier	PT
HL -INT	Interviewer tone tier	PT
HLF	Pragmatic function	IT
Misc	Miscellaneous tier	IT
Com	Comments tier	IT
React	Interactions tier	IT
Topic	Topic tier	IT

IT=Interval Tier, **PT**=Point Tier

Each tier will be explained step by step:

Anonymization and Data Protection

Names, addresses, and phone numbers should always be anonymized. Usually there is a procedure for this steps when the recordings were transcribed. However, something can always be missed. Some recordings also blend this kind of information out by playing a beeping tone instead of the name, addresses, or phone numbers of the speakers. Should a name not have been anonymized you can annotate either <name>, <address>, and <number> instead of the actual private information.

Sample Annotations

An example of what the (final) annotation should look like the representation in the Figure 2 (see Section 3.5). It is an A0 *TextGrid*. The annotation process is supported by the waveform (top) and the spectrogram (middle) and the pitch track (blue line within the spectrogram). The pitch settings in this example are set to 100 Hz to 400 Hz for the female voice of speaker BE018, a young speaker of British English. Below the spectrograph are the 11 manually annotated tiers (from top to bottom): tasks tier (Task), speech acts (SA), intonation phrase (IP) tier, intermediate phrase (ip) tier, tone tier (HL), pragmatic function tier (HLF), interviewer tone tier (HL - INT), the miscellaneous tier (Misc), the comments tier (Com), the interaction tier (React), and the topic tier (Topic). Under the manually annotated tier, you can see the duration of the utterances. Under that you have the option to zoom in and out on the speech samples.

Step 1. Before starting to annotate the recordings:

- Choose 45 recordings (45 different speakers) from each group (e.g. LINDSEI-CZ, LOCNEC, etc.). Try to find a balanced amount of female and male speakers if possible. However, sound quality should always be the primary criterion for choosing a recording. Listen to the recording to see if it can be used. If the audio quality is too low, it should not be annotated.
- See if the recording is in a wav format (16-bit and not 24-bit). It should not be transcribed as an MP3 file since the time line can be off by several milliseconds and this could change the results.
- Look at the transcript of the recording and set a limit of 550-580 words if the speech begins with the monologic part of the interview. If it starts with the picture description, then select the first 550-580 words after the picture description.
- Please do not delete the transcriptions that follow after the 550 words that you have transcribed. Simply leave a line break. That way we can easily add words if some passages have to be deleted. In the last version of a transcript there will only be the words included that were actually annotated in the *TextGrid*.
- The beginning of a recording usually starts with the setting-up of the scene and speech by the interviewer. You can completely ignore this part and start annotating where you can clearly hear the first complete utterances by speaker B (i.e. the beginning of the monologic part). However, try to annotate passages that are uttered a little bit later in the interview, since some students might be too nervous and need some time to adjust to the recording situation. For some speakers this might not be the case and the passage at the beginning might be more “natural” (sometimes even better than later in the recording). Your choice depends on the situation.
- Transcribe material first that is in a good audio quality and which ideally starts with a clear monologic part. Dialogues (a lot of interviewer - interviewee interaction) will be chosen last once all monologues have been transcribed.
- Transcript: Delete all annotations from the annotated 550-580 words of speaker B (the interviewee) and all interviewer stretches (if present). These annotations were deleted because we had more accurate or other ways to mark these in *Praat*. For example, We

deleted annotations for pauses because pauses are clearly marked by time stamps and are more accurate then the very generic “., .., …” used in the LINDSEI and LOCNEC annotations. The following annotations were deleted:

. short pause

.. longer pause

… very long pause

: which is attached to the word to signal syllable lengthening

- The following brackets can be deleted : <…> …</…>&<...> e.g.:
    - <overlap> … </overlap> --> content can also be deleted since we do not want overlapping speech
    - <A>...</A> the interviewer’s speech can be entirely deleted
    - <foreign>… </foreign> --> the foreign word can stand by itself without the annotations. Foreign words are sometimes annotated as <foreign>…<foreign> or in between §...§. Please delete all §...§.
    - <B> … </B> --> the content should remain, only the annotations should be deleted
    - <laughs> --> can be deleted entirely
    - the[i:] --> [i:] should be deleted
    - Only “=” (for incomplete words) and “<X>” “<XX>” “<XXX>” (for one or more unclear words, each X standing for a word) are left for our purposes.
- **Filled Pauses:** Note that there can be different transcriptions for filled pauses: *er*, *ehm*, *eh*, *uhm*, *uh*, *mm*, *mmh*, etc. These filled pauses should be standardized and transcribed as *er*, *mm*, and *um*. Additionally, in some transcriptions the filled pauses are transcribed in brackets (*er*), (*erm*), etc. when you replace them with *er*, *mm*, and *um*, make sure to get rid of the brackets as well. You can use the replace function in Excel or Notepad++ to replace these filled pauses by the standardized filled pauses: i.e.: *er*, *mm*, and *um*. Be careful when you use the search and replace function so that you do not replace parts of words such as in the word “singER”, where the last “er” would be deleted as well.

Step 2. Opening files in Praat and adjusting the pitch settings

- Open the *Praat* Objects window
- Click on ‘Open’ → ‘Open long sound file…’ or use Ctrl+L/Strg+L
- Upload the recording and the corresponding *TextGrid* file that you have previously worked on.
- If no *TextGrid* file has been created yet and you are working with a new recording, then click on the right hand side on ‘annotate’ →’To TextGrid…’
- Then a *TextGrid* appears in the *Praat* Objects Window and you have to select both and click on ‘View & Edit’
- The program then asks you to enter all interval tiers and to specify which ones are point tiers. Delete the default names and enter *IP ip HL Misc Com* into the first row

for interval tiers. Then add *HL* as a point tier and click on apply. Each entry should be separated by a space. Only five tiers at a time are added for visual purposes. The other tiers specified at the beginning of this annotation manual were added by the project coordinator (Karin Puga) at a later point.

- Adjust the *pitch settings* according to the gender and voice of the speaker: The standard pitch settings in *Praat* are set at 75 Hz to 500 Hz. Adjust the advanced pitch settings to 100-400 Hz and change the representation of the f0 to speckles. If the voice is too deep usually with male speakers then set to 50-400 Hz.
- To remove a boundary line on the tiers you can either select ‘Boundary’ in the top tab and click on ‘remove’ or you can select the boundary and press Alt+Backspace/Leertaste.
- To save a file in *Praat* click on ‘File’ on the top tab and select ‘SaveTextGrid as a text file ...’ or press Ctrl+S (Strg+S). Always remember to save the *TextGrid* while annotating. This does not happen automatically and there is no auto backup copy. The software can break down occasionally and it would be a waste of time to forget saving the file multiple times while annotating.

Step 3. Task tier

In each LINDSEI and LOCNEC recording there are three different tasks each speaker has to fulfill:

Task 1: monologic part about one of three topics (country, movie, experience)

Task 2: informal interview about different topics (e.g. course of studies, family, etc.)

Task 3: picture description

On the Task Tier, only 2 annotations are to be annotated, i.e. either “1” (for monologues), and “2” (for the informal interview). In Figure 51, you can see that the intervals are either filled with 1 or 2. The picture description (Task 3) was not annotated on any other tiers and will not be included in this tier either.

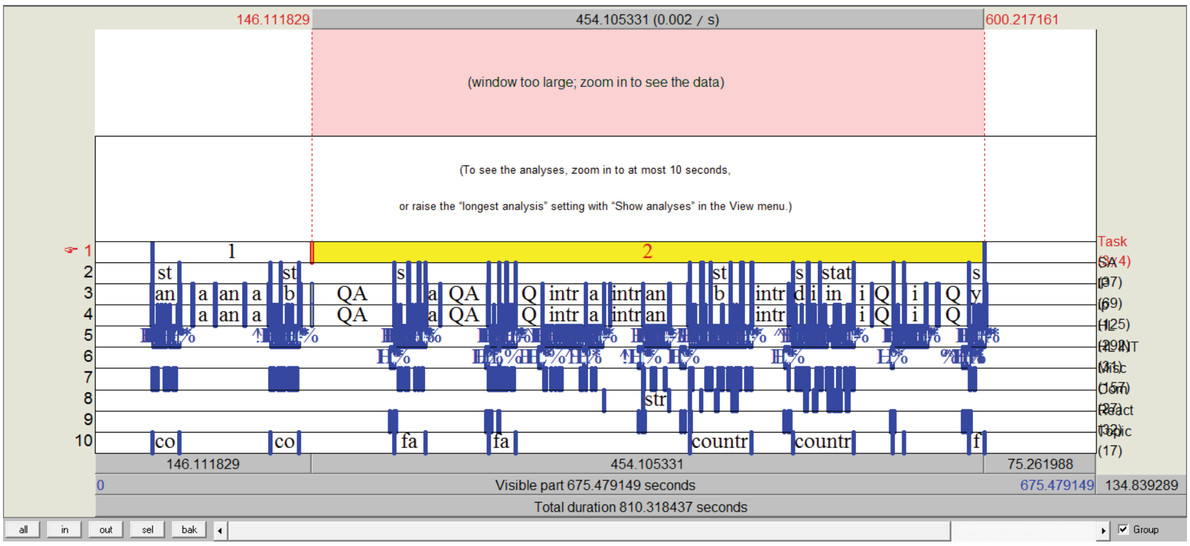


Figure 51. *Praat* screenshot of task tier annotations (file GE001)

Sometimes the boundaries are not clear between Task 1 and 2. In most cases you will have to decide which task would be more characteristic of the speech sample. Generally, while the interviewee is speaking about his/her topic of her choice without being interrupted, this would be annotated as Task 1 (monologue). Once there is a question in between and it continues as a dialogue, even though the topic remains the same, this would count as Task 2 (dialogue). In this case the Task 2 annotation would start with the interviewer's first question. If the interviewee's speech is only interrupted once briefly or the question is ignored altogether and the part continues as a monologue for a longer period of time, then this would be annotated as Task 1 (monologue) still.

Step 4. Speech acts

It is important to differentiate between different speech acts that are used in the recorded utterances, because they fulfill different functions and can be accompanied by different prosodic patterns.

Speech Acts that can be distinguished on interviewee tiers:

- **Statements**
- **Answers**

However, all questions by the interviewee (however rare) were taken out and therefore only statements/declaratives were included in the corpus.

Answers (A): Task 2 (the informal interview) would most commonly have speech acts labeled as "answers". Once an interviewer asks a question all subsequent utterances from the interviewee are labeled as "answers". However, there are sometimes questions posed by the interviewer that resemble more statements but are in fact questions. In these cases the interviewee utterances were labeled as "answers" (see example 12).

(12) BE010: interviewer question: "But I guess it takes quite a lot of time to write things for that" - interviewee answer: "it depends sometimes um you sit there and you think no there's nothing going on in here"

Statements (S): Utterances were classified as "statements" that are usually part of Task 1 (monologue) and that are mostly free of any influence from the interviewer. When the interviewer makes a statement, the following utterances by the interviewee are labeled as "statement". Once a question has been posed by the interviewer, the subsequent utterances are usually labeled as "answers", however, when the interviewee starts off with a new topic all by himself/herself after a while, these utterances are labeled as "statements".

Step 5. Intonation phrases

Intonational phrasing

The delimitation of IU boundaries is a quite difficult task, especially spontaneous speech, which contains hesitations, repetitions, grammatically incomplete sentences, and which depends on different contextual factors (cf. Gut 2009: 208). Even though a speaker is generally free in choosing the length of an IU, there are certain characteristics of IU frequencies and -lengths in native English speech. One such characteristic is for instance that intonation phrases mostly coincide with clause boundaries and ips sometimes mark subordinate clause structures (cf. Gut 2009: 209).

Internal criteria, i.e. phonetic cues present within an intermediate or intonation phrase

- An IU includes at least one pitch accent on an accented syllable and edge tones (i.e. a phrase accent and a boundary tone which may appear on unaccented syllables).
- Therefore, short utterances comprising only unstressed syllables are thus excluded from the annotation procedure or attached to previous IUs.

External criteria, i.e. phonetic cues present at the phrase boundary for the identification of intonation phrase boundaries (cf. Cruttenden 1997):

Final syllable lengthening

- Anacrusis (i.e. an accelerated production of the initial unstressed syllables)
- An intonation boundary tone (see table in Step 6)
- Change in pitch level and direction of unaccented syllables (F0 resetting)
- Pauses (pauses may be longer at IP boundaries than at ip boundaries cf. Keseling 1992)
- Not every pause is produced to mark an intonation phrase boundary
- Occur in hesitations, before restarts and repairs
- Biological necessity of breathing

Annotation steps for intonation phrases (IPs):

- Listen to the recording and segment the utterances into larger IPs by placing interval landmarks (time-stamps).
- Copy and paste the utterance within the interval from the respective transcript into the larger intonation phrase. You can find the transcript in the RTF file that you have prepared in Step 1.
- Transcriptions of cliticizations such as *can't* are possible.
- Some transcripts have different annotation conventions. For instance, some include '…' for pauses. Delete those and remember that these types of annotations do not count as words into the final word count. Other annotations are also deleted if you have not done so in the first step.
- If you hear something different in the recording than it was transcribed you can change it during the annotation process, however, always write a comment in the Com Tier when you make changes, especially if you are unsure.
- Generally you can ignore long question sessions by the interviewer and short answers such as *yes*, *no*, *okay* by the interviewee because they do not represent full IPs and are simply too short to be analyzed from a prosodic perspective.
- IPs should ideally be "completed". By completed, I do not mean syntactically complete. They should rather fulfill the internal criteria for IPs mentioned above. According to the ToBI annotation system, there has to be a certain structure that

makes up an intonation phrase. Each intonation phrase consists of one or more ips. Each ip needs to include at least one pitch accent and one phrase accent. The last ip simultaneously ends with the intonation phrase, therefore, there will be one pitch accent, one phrase accent and one boundary tone. This internal structure has to be given to be counted as one completed intonation phrase. If the last ip is for example interrupted with a laugh or by the interviewer, then the entire utterance will be discarded.

- There are some exceptions within ips in longer IPs, if there is one ip that was taken out in between, and then this does not mean that the entire IP will be taken out. As long as most of the intonation phrase is "complete" and there is a clear phrase and boundary tone at the very end, the intonation phrase can be left in the annotations.

Step 6. Intermediate phrases

The same procedure will be carried out for the intermediate phrases (ips) as for the IPs in Step 5, just that you do not need to select 550-580 words anymore and speech phenomena that were already annotated for the intonation phrase tier can generally be copied to the ip tier. For example, if there is a longer pause on the intonation phrase tier the same time stamp will be copied into the ip tier.

Often it is difficult to determine ip boundaries because phonetically many words are getting linked and cannot always be clearly separated.

Also insertions or sudden changes of the chain of thought that are indicated by discourse markers like "I mean" or "you know" are sometimes annotated as separate ips and sometimes as one ip depending on the context.

|| the two lines represent intonation phrase boundaries

| one line represents ip boundaries

This will not be used in the actual annotations but serves for demonstration purposes only. For example, the following sentence would be treated differently:

|| I was there yesterday *I mean* today ||	IP tier
|| I was there yesterday I mean today ||	ip tier
| hp |	Misc tier

Explanation: The sentence is treated as one ip but the self-correction is marked as "hp" in the Misc Tier. **Discourse markers** are not always treated as part of an ip, as can be seen in the following example:

|| I was about to … to …you know … change my clothes ||	IP tier
|| I was about to to | you know | change my clothes ||	ip tier

Explanation: in this made up example the discourse marker "you know" is clearly separated by pauses and it would therefore be counted as one ip.

Incomplete phrases are usually also counted as separate ips:

‖ I was about to … oh you know we went to the gardens ‖	IP tier
‖ I was about to | oh | you know we went to the gardens‖	ip tier
| inc |	Misc tier

Explanation: this is an example of a sudden change of thought. The first sentence would be marked as incomplete, "oh" would constitute a separate ip depending on the emphasis/tune it receives and "you know" would be part the next ip but also depending on its tonal patterns. If "you know" was followed by a longer pause, then it would be annotated as a separate ip.

Similarly difficult is the treatment of **hesitations or self-corrections**. Sometimes annotators tend to annotate repetitions as separate ips. This also depends on the context and cannot be generalized. Mostly hesitation phenomena would be marked as one ip and annotating the hesitating part in the Misc tier. In other situations they could be separate ips.

Sometimes it might be tempting to annotate phrases that syntactically and topic-wise belong together, but in spontaneous speech this might not always be the case for ips. Even longer breaks in between the sentences do not always indicate a separation of ips. In these cases the tone should primarily be taken to indicate an IU boundary.

Lists: When interviewees list things, these items were generally treated as separate ips, e.g. ‖ I really like bananas | apples | and cherries ‖ unless the intonational patterns did not separate these items from each other clearly enough.

To be able to tell how long each ip is, one needs to distinguish between silent pauses and breathing pauses:

up	unfilled pause
b	breath

These are annotated in the interval tiers within the IP or ip phrase tier. Sometimes pauses are only marked in the Misc Tier. This will be explained further in the sections below.

Example:

The utterance "I went to the kitchen and made me a sandwich" can be either uttered as one intonation phrase or two or more ips, depending on the context, intonation, and pausing behavior (among other factors).

Intonation Phrase Tier: ‖ up | I went to the kitchen and made me a sandwich |int ‖

Intermediate Phrase Tier: ‖ up | I went to the kitchen | up | and made a sandwich | int ‖

The two ips do not necessarily have to be separated by a pause but the boundaries can also be marked by a different tonal patterns.

Filled pauses

Filled pauses by themselves are never annotated as separate ips. If there are too many filled pauses within an utterance this may be transcribed as ‘hp’ (hesitation phenomena). Other annotations of filled pauses are explained in Step 9.

Step 7. Tone Tier

On the tone tier, three different categories of tones are annotated:

Pitch accents	characteristic pitch heights (H or L) which are associated with prominent syllables within an ip
Phrase accents	pitch height (H or L) at the end of ips
Boundary tones	pitch height (H or L) at the end of utterances of ips

There can only be high (H) or low (L) tones. H and L targets are taken as relative terms and do not indicate a rise or fall, respectively, but can also indicate a flat, level, or non-rising f0 (cf. Warren 2016: 12). Following Patterson’s (2000: 67-68) approach, f0 measurement points selection were taken at local highs and lows. A high pitch (H) was coded if there was a clear local f0 peak. A low pitch (L) was coded for local f0 minima in an intonation phrase. The following tones are allowed:

H	high tone
!H	downstepped high tone (replaces H*+L)
^H	upstepped high tone (replaces L+H*)
L	low tone

These tones can appear in different positions:

% H, H/L*****, H/L**-**, or H/L**%**   / = either H or L

Only pitch accents can be annotated as combinations of tones, which are marked with a **+**

Should the falling movement start within the target syllable, then it is to be labeled with an H*+L accent, otherwise if there is no significant change in f0 height there, then the token is to be treated as carrying a simple high-level tone (H*).

Place point annotation exactly on f0 contour to be extracted later. Make effort to avoid choosing values that appeared to be due to segmental perturbations of f0 (also called “microprosody”; cf. Ladd 2008: 25). In these cases choose the last f0 value before the beginning of the microsporosdic dip. For IU final lows choose the lowest reliable f0 value.

In order to obtain the f0 measurements, the 550 words long samples for each of the 225 speakers were annotated according to ToBI labels (see Section 3.5). In order to decide whether a high (H) tone or (L) tone would be annotated, the F0 tracking (called pitch tracking in *Praat*) was used in *Praat*. Following the *Praat* recommendations, the pitch settings were changed to 100 Hz “pitch floor” and 500 Hz “pitch ceiling” for female voices. However, in some cases there were female participants who had a rather male voice and vice versa. In these cases the pitch settings were changed to the opposite sex. Cases such as creaky voice or octave errors or spurious f0 values in voiceless parts of the signal complicate the f0 tracking. A pitch floor setting of 100 Hz excludes most instances of creaky voice. Therefore, every single IP was carefully inspected by the help of the automatic f0 tracking, auditory verification, and visual inspection. The ToBI labels were thus placed at a point in the f0 contour where the point would the highest (H) and/or lowest (L), in that way only where the point is placed in *Praat* will the *Praat* script later measure the exact pitch height in Hz. Thus, the linguistic labels on the point tier (Tier 5) were only placed on inspected f0 contours which match the auditory verification and therefore excluding f0 artifacts.

Pitch accents

Pitch accent tones are marked at every accented syllable. If a pitch accent is not assigned to a syllable this means that the syllable is not accented. Each ip needs at least one accented syllable.

Pitch accents are marked by a star ‘*’ and the following pitch accents are to be annotated:

H* high pitch accent - an apparent tone target on the accented syllable which is in the upper part of the speaker’s pitch range for the phrase. This includes tones in the middle of the pitch range, but precludes very low F0 targets.

!H* downstepped high pitch accent on the accented syllable, which is in the upper part of the speaker’s pitch range for the phrase. Note also that this diacritic is NEVER applied to the first H tone in a phrase. This tone can only appear after at least one high pitch accent has appeared in the ip. Downstepped means that the tone is lower than the preceding high tone(s) due to declination in speech but still not low enough to be marked as a low accent (cf. Cruttenden 1997: 60; Gut 2009: 199).

^H* upstepped high pitch accent realized on the accented syllable, which is in the upper part of the speaker’s pitch range for the phrase. It is a relatively sharp rise. This tone can appear anywhere within ips as long as it can be clearly distinguished from a regular high pitch accent.

L* the low pitch accent is an apparent tone target on the accented syllable which is in the lowest part of the speaker’s pitch range.

There are four simple pitch accents as described above and three compound ones, which are marked by a + linking the two tones, e.g. H+L*.

Combinations of low and high pitch accent when the low tone is accented

L*+H a low tone target on the accented syllable which is immediately followed by relatively sharp high tonal target in the upper part of the speaker’s pitch range.

H+L* a low tone target on the accented syllable which is immediately preceded by a relatively sharp high tonal target in the upper part of the speaker's pitch range.

L+H* a high pitch accent on an accented syllable which is immediately preceded by a marked low level on an unaccented syllable.

Explanation to phrase accents and boundary tones:

Phrase accents and boundary tones can only consist of a single tone: either H or L targets (see also Gut 2009: 200).

Phrase accents stretch from the last pitch accent of an ip to the beginning of the following ip or to the end of the IP.

The boundary tones fall exactly on the IP boundary.

Since the end of an IP is also the end of an ip, there are four combinations of phrase accents and boundary tones that are possible at the end of an IP: L-L%, H-L%, L-H%, and H-H% (cf. Warren 2016: 11-12).

Edge tones in English are interpreted as follows:

- H- and H% signals cohesion between the current phrase and the subsequent phrase.
- L- and L% signals separation of the current phrase from the following one.

Phrase accents

All phrase accents occur at an ip boundary.

The last tone marking the end of an ip is marked by the diacritic "-".

The following four phrase accents are possible:

H- high phrase accent

!H- downstepped high phrase accent

^H- upstepped high phrase accent

L- low phrase accent

Boundary tones

All boundary tones occur at IP boundaries.

The very first or last tone marking the end of an IP is marked by "%".

%H initial boundary tone, which occurs at the very beginning of IPs. %H marks a phrase that begins relatively high in the speaker's pitch range; the default initial boundary is in the middle of the range or lower, and will be left unmarked in the annotation. %H should be used only when a high pitch at the beginning of an IP cannot be attributed to a H accent, i.e. when the first word itself is not accented.

All other boundary tones occur at every full and final intonation phrase boundary.

H%	high final boundary tone
!H%	downstepped final boundary tone
^H%	upstepped final boundary tone
L%	low final boundary tone

Uncertainty

On the tonal tier, two kinds of uncertainty may be indicated: uncertainty over whether an event of a particular type has occurred and uncertainty over the tonal value of an event that clearly has occurred. You may be unsure whether a particular syllable is accented, or, knowing that it is accented, you may be uncertain of the accent type.

Uncertainty of the first sort (whether the event has occurred) is indicated by *?, -?, and %? for pitch accents, phrase accents, and boundary tones, respectively.

Uncertainty of the second sort (over the tonal value of a clearly occurring event) is indicated by *, -, and %. Thus, for example:

* means “This syllable is definitely accented but I am not sure which accent type should be assigned”.

*? means “I’m not sure whether this syllable is accented or not”.

The same symbol can be applied on the pitch tier (tier 3), i.e. L?, H? - if you are uncertain about the pitch level that is used on this tier. An empty point tier indicates that you are uncertain if a pitch event exists in this example.

Step 8. Pragmatic function tier

Intonational meaning has multiple layers (e.g. linguistic or lexical meaning; discourse, interactional or social meaning; paralinguistic or attitudinal meaning) and a single intonation contour can convey meanings simultaneously at more than one of these layers (Warren 2016: 66). There are also descriptions of uptalk that reveal multiple functions or interpretations within one layer of meaning (Warren 2016: 66). The pragmatic function tier includes the functions of the speech acts uttered in the annotated stretches.

Pragmatic functions:

Generally, the following tone functions were distinguished on the ip level:

- checking, continuation, emphasis, finality, handing over of turn, and insecurity.
- H-tunes are generally associated with: continuation, checking, insecurity, and listing
- L-tunes are generally associated with: continuation, finality, handing over the turn to interlocutor, emphasis

Continuation: Most common pragmatic function in the corpus. Commonly used in monologic parts and longer answer sessions, in which the speaker wishes to continue holding the floor. Different from the "checking" function described below. They are not formulated to ask questions and they are typically followed by another speaker turn by the same speaker. These continuation turns are followed by short or no pauses and are usually not interrupted by interviewers. However, in some cases they are accompanied by uptalk tunes which simultaneously ask an indirect question along the lines of "I am going to continue talking/holding the floor, is that ok? or are you with me?"[86] These types of continuation utterances are at times difficult to distinguish from checking utterances. Continuation may also be used as a preview of a longer speaker-turn or -turns. That way the listener is signaled to prepare him-/herself for a longer run and might be expected to backchannel the utterances at turn breaks to keep the interaction going or "permitting" the interviewee to continue. Previous studies have labeled utterances fulfilling this pragmatic function as "floor holds" (e.g. Ritchart and Arvaniti 2014; Arvaniti and Atkins 2016).

Handing over turn: The speaker intends to cede the floor and hands the turn over to the interviewer because he/she runs out of things to say or would like the interviewer to contribute to the conversation as well. Generally, these type of utterances are accompanied by low tones and a softer or whispery voice and a longer pause followed by interviewer speech.

Listing: Listing utterances have a particular melody and involve the listing of several items. The items listed are often separated into smaller chunks (ips) and typically accompanied by high phrase accents on each item in the list and a final low boundary tone. These utterances are very common in both monologic and dialogic parts. Rarely these type of utterances are accompanied by uptalk tunes and are indirect questions, which ask whether the speaker should continue listing items. Often the interviewer asks questions like these, sometimes in a more explicit way, such as asking whether they have said enough and should stop speaking. These types of questions were excluded and marked as "asides".

Finality: These type of utterances are very similar to "handing over turn" utterances. They also occur at the end of monologic stretches and in shorter question and answer sessions. However, they are not uttered in a softer voice but with more certainty and determination. Utterances signaling finality are produced with a more extensive and clearer low boundary than "handing over turn" utterances.

Checking: Checking involves ips that are formulated as indirect questions to negotiate common ground with the interlocutor, i.e. to elicit feedback/backchanneling from the listener. Also used to give the listener an opportunity to add something or to ask questions. What the speaker is asking for in these type of utterances is if the listener is still paying attention, agrees with or accepts what has been said, or just simply to check comprehension. Minimal responses from the listener help to identify these types of utterances. Very often these types of utterances are followed by a rising intonation pattern and strongly coincides with uptalk tunes. In other studies, this pragmatic function is also called "confirmation request" (e.g. Ritchart and Arvaniti 2014; Arvaniti and Atkins 2016).

---

[86]These are constructed examples of what the uptalk tunes may mean on the pragmatic level. Actual examples from the present data set are illustrated in the discussion section.

Insecurity: This pragmatic function mainly found in non-native speech involves the use of high tones that is atypical in comparison to the speech signal in other parts of the sample. This function is similar to the "checking" function, in that it sometimes asks through the use of high tones whether something was understood or correct. It is also accompanied by uptalk instances in some cases.

Emphasis: This refers to ips that include (exaggerated) emphatic speech and/or narrowed focus in marked declaratives.

Step 9. Miscellaneous tier

Whenever a stretch should not be annotated or whenever a speech or non-speech event occurs that influences the interviewee's speech, you can use the following abbreviations in the miscellaneous tier. The interviewer's speech is mostly ignored such as other incomplete sentences for various reasons. However, set an interval for each of these instances within miscellaneous tier intervals and add an abbreviation for those instances. How to transcribe passages that are excluded, see the list below:

Speech events

There are different speech events that are distinguished on the *misc tier*:

as asides, i.e. comments on the recording situation

bc backchanneling (exclamations by the interviewer such as *mhm*, *okay*, *right*, etc.)

cv creaky voice

e elongated phoneme (represented by ":" in the transcriptions, e.g. *yes*:) consonant or vowel lengthening

fp filled pause (used if filled pauses are not annotated as "e" or "hp")

fr foreign words

hp hesitation phenomenon (self-corrections, repetitions (not for emphasis), stuttering, filled pauses followed by silent pauses or multiple filled pauses, elongations included)

imit imitation (i.e. interviewee imitating another speaker or herself/himself) also including indirect speech "oh and then he said…")

inc incomplete sentences

int speech by the interviewer

intr interruptions by the interviewer or other external sources (e.g. noise from open window)

o	overlapping speech
QA	question and answer-sessions interviewer + interviewee, and short answers given by the interviewee
unc	unclear stretches of speech
wb	word break (when there is a longer silent pause between a word and the final consonant)
whis	whispered speech

Non-speech events

The following non-speech events are annotated:

ant	antrophonics, e.g. cough, laugh, click of tongue, etc.
b	breathing pause
bn	or other background noises, signal interfering (probably cell phone)
br	break in recording, transmission problems
up	unfilled pauses or silent pauses are only annotated on the Misc tier if there is a clear pause but still no intermediate or intonation phrase boundary is perceived

- Make sure to use the lowercase characters with the codes as they appear in this manual. For example, do not use HP but hp.
- If the interviewer interrupts the interviewee in between a silent pause this is marked as “int”. Only if the speaker stops his/her speaking flow then the entire stretch is marked as “intr”. If the interviewer interrupts within the speaking flow that the interviewee continues this is marked as “o”.
- If someone whispers, has a creaky voice or utters a word or a passage in a foreign language the tonal tier (HL Tier) usually remains untouched. It is difficult to judge from the f0 line in *Praat* and it does not make sense to annotate the tonal tier for these passages. Still in some cases it makes sense to annotate IUs anyways, even if they are creaky.
- When a speech event or non-speech event occurs that is not transcribed it will always be marked in the *misc tier*. However, if a gap occurs in the IP or ip tiers it will still be annotated for those as well. Even if this sometimes means that there will be, for instance, “int” on two tiers (IP, ip).
- The speech event “hp” can be any kind of hesitation phenomenon, hesitations (e.g. er … er … um, th=the, often this kind of hesitation phenomena also include silent pauses in between), repetitions (e.g. I IIII have seen him), and self-corrections (e.g. I think … er I don’t think). When a repetition or self-correction occurs: only mark the first word or couple of words before the corrected and last word. Do not include the last or corrected word in the hp-interval.

- Whenever an unfilled pause would occur just by itself it will be clearly marked as "up", however, if it occurs together with filled pauses and other hesitation phenomena like repetitions or stuttering or self-corrections then these stretches are taken together and marked as "hp" and are not distinguished as a sequence of many phenomena.
- Sometimes long stretches of speech that are marked with "intr", "ant", or "hp", do not mean that the entire stretches were interrupted, laughed, coughed, or repeated. Typically, the end of a sentence was interrupted and the sentences were therefore not completed or the speaker laughs while uttering certain words and therefore the entire stretch was marked as "ant".

Examples:

Intermediate phrase Tier: | w= what is that |

Misc Tier: |hp |

Intermediate phrase Tier: | I don't think they they will come |

Misc Tier: | hp|

What if two incidents happen at once, e.g. "int" and "whis"?

You should decide for the one that is more prominent. There should not be any double annotations on the misc tier or multiple misc tiers. If there is something equally important that has to be annotated too, just transfer it to the com tier.

How to deal with overlapping speech?

Only speech by the interviewee which is completed will be annotated. However, overlapping speech or background noises that do not interfere with the interviewee's speech, these passages can still be annotated but these events are indicated anyways on the miscellaneous tier. For example, if there is a phone ringing in the background but the interviewee does not stop talking and the ringing of the phone does not interfere with the f0, then the passages will still be annotated and the interference will be annotated on the miscellaneous tier.

Example:

For example, if an utterance by a speaker begins, but is then interrupted by an interviewer it can be annotated as follows on the intonation phrase tier:

<A>I just wanted to <overlap>say that

<B> I can't</overlap>

Miscellaneous Tier: | inc| o |

If the speech does not overlap it can be transcribed as: | inc| intr |

Step 10. Comments tier

- This tier can be used to leave any kind of comments for the next reviewer of the annotated file.
- This tier will be left in, since other Misc Tier annotations that did not fit into the Misc Tier, were transferred to this tier. Therefore, this tier also includes important information that happens simultaneously with another event on the Misc Tier.

Step 11. Reaction tier & HL - INT

The *react tier* includes all interactions between the interviewer and interviewee which have a direct influence on the annotated IPs.

On the "React" Tier only three annotations are possible:

interviewer question	the question that is asked by the interviewee that has a direct effect on the interviewee's speech (this can include wh-questions, yes-no-questions, echo questions, intonational questions, these were, however, not differentiated)
interviewer statement	the statement or comment that is made by the interviewee that has a direct effect on the interviewee's speech
interviewee reaction	the answer that is given by the interviewee to the interviewer question or statement

The *react tier* annotations will only be included if there is a direct influence on the statements that were annotated. As can be seen in Figure 52, there might have been more question and answer pairs in the QA section that was annotated on the IP and ip tier. However, only the last interviewer question was annotated which had an effect on the following statement. So if the statements received an interval on the SA tier, this means that you should check whether there was an influence by the interviewer immediately before. The interviewee reaction usually includes the beginning of the following statement/intonation phrase, however, the reaction can start before that and continue along those lines, just as in the example below. Each question receives one separate "interviewer question" annotation. If questions are ignored by the interviewee, you do not need to annotate anything on the *react tier*. The interviewee reaction should not be dragged until the end of the intonation phrase but rather just maybe about the length of the first ip. The interviewer question/statement annotation should be annotated accurately and depict the length of the actual question or statement.

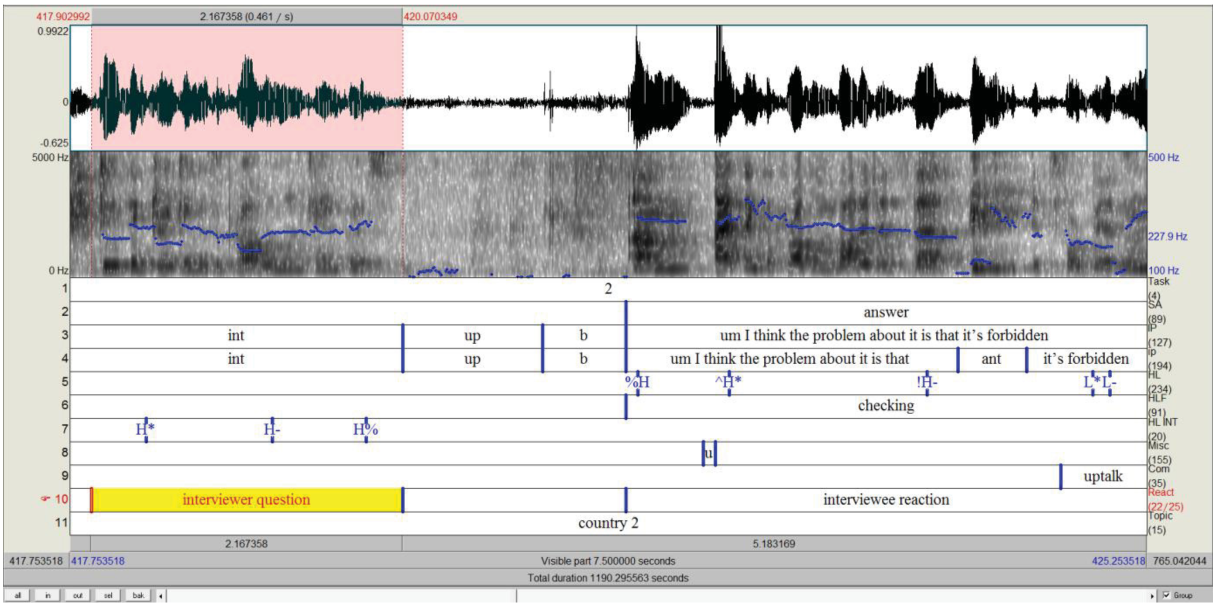


Figure 52. *Praat* screenshot of react tier & HL-INT annotations (file GE028)

HL - INT

Just as in the figure above, the annotated interviewer stretches on the *react tier* receive tone annotations on the HL- INT tiers. There are no segmentations into ips but they should be insinuated by the tones on the HL - INT Tier.

Step 12. Topic shift tier

After having annotated the *react tier* with the interactions, it will be now easier for you to determine topic shifts. Topic shifts are usually initiated by interviewer questions. There is a list of possible topics below. If you encounter a topic which does not fit into any of the topics below, you can add a new category. But please be careful not to add terms which include similar topics. Anything that has been excluded from the annotation, such as "intr" or "QA" sessions, these do not receive a topic annotation. Only utterances by the interviewee receive a topic annotation.

List of detailed topics:

country if they speak about a country they have been to, including the food, architecture, sightseeing, the people in that country, etc.

country 2 if they talk about another country or countries.

studies if they talk about their course of studies or school

food if they talk about food and cooking and it doesn't have anything to do with the country they talked about previously

politics if they talk about politics and it doesn't have anything to do with the country they talked about previously

jobs if they speak about jobs that they have/had or would like to have

future plans	if they talk about future plans such as moving to another place, etc.
experience	if they talk about an experience that does not fit into any of the other categories
family/friends	if they talk about family history or talk about people close to them, friends, roommates, colleagues, boyfriends/girlfriends, etc.
leisure	if they talk about hobbies or other free-time activities they have done or do regularly
play/movie	if they retell the plot of a play, a movie or a series
book	if they talk about a book
self	if they talk about themselves, their character, do a self-evaluation, count their weaknesses and strengths

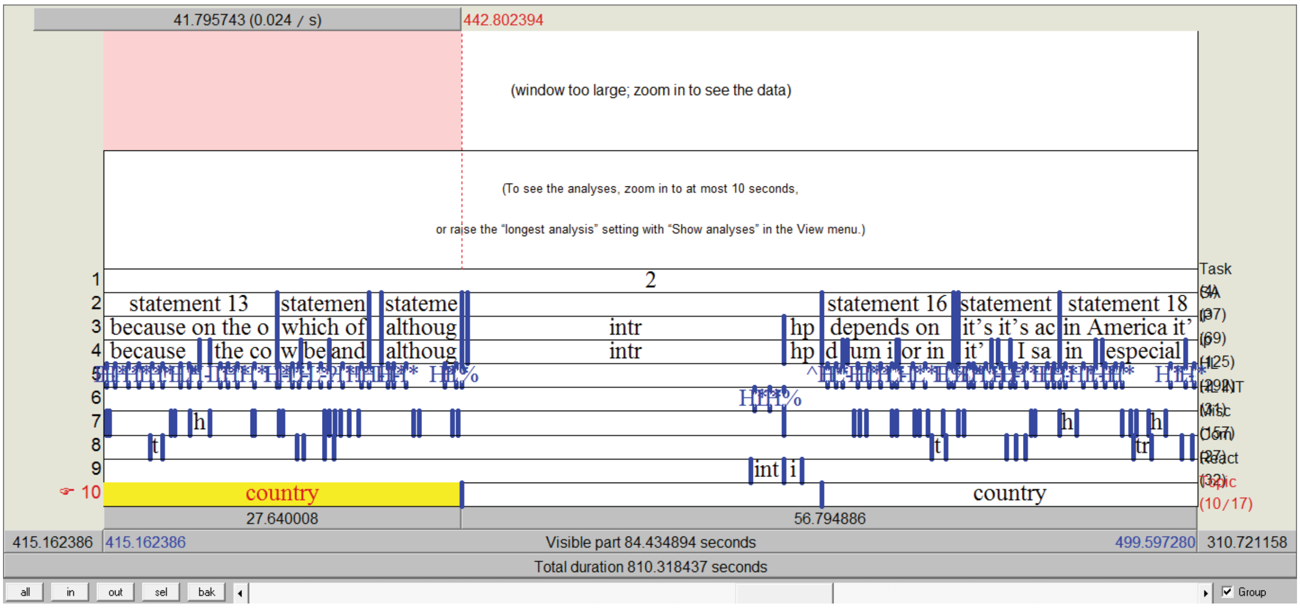


Figure 53. *Praat* screenshot of topic shifts annotations (file GE001)

As you can see in Figure 53, the “intr” stretch that was excluded here, does not receive a topic annotation. Also the pauses right after and before the statements were also not considered. Pauses in between IPs that are not interrupted are however counted as one topic.

Step 13. Annotation Checks

There are **three major steps**, which have to be completed for every single recording:

A0 First annotation of the recording, including steps 1-8 in this manual

A1 First annotation check by a different annotator

A2 Second and "last" annotation check by another annotator who was not involved neither in A0 nor A1.

Check in the overview file in our Dropbox folder which *TextGrid* file that has to be checked next.

Listen to the entire annotated stretch and move boundaries (time-stamps) or point tiers around where you see fit. Also pay special attention to the comments tier (Com Tier 8) since this tier might include directions or problems by other annotators.

A1

- **Metadata-base:** Make sure your name is properly marked in the metadata-base. Please check if all the information about the recording and the interviewee and interviewer has been entered properly. Enter any missing information.
- **Tiers:** Check whether the tiers have been named correctly and whether they are in the right order (Task, SA, IP, ip, HL, HL-INT, Misc, Com, React, Topic). If you disagree with an IP, ip, or HL then add a second tier under the ones made by A0 respectively, then name your tier just like the first one but with your initials. So in A0 the first tier would be "IP" and in A1 you would add "IP - KP" (standing for intonation phrase tier by Karin Puga) right under the first IP made by A0.
- **Annotation Conventions**
    - Check whether the annotation symbols have been deleted (only "=" (for incomplete words), "<X>" "<XX>" "<XXX>" (for one or more unclear words, each X standing for a word), are allowed) -everything else should be deleted. Especially ":" and "<?>" are mostly overseen.
    - Check whether all *erm, em, ehm, uhm* have been replaced with *um*, and all *eh* with *er*, and all *mhm* with *mm*.
    - Delete all double spaces in the text file as well as the *TextGrid* file.
    - Correct these types of mistakes in all IUs, since you do not know which option the last annotator will decide for.
- **Intonation phrase tier (IP)**
    - If you see that less data has been transcribed, and then please annotate more, each file should have at least 550-580 words (if possible). If more than 590 words have been annotated, then please delete a few passages, preferably at the beginning of the file, or at the end but never in between. Put the additional IPs into your tier with your initials.
    - If you disagree with an entire intonation phrase then put your version in the tier that has been marked with your initials. If you would not change any IPs and you totally agree with all of them, then you do not need to add a second IP tier with your initials. That way we know that both A0 and A1 agree with each other regarding all IPs. Only add annotations to your tier that you disagree with. If you agree with an intonation phrase, for example, then you leave this part on the second tier blank.
    - Make sure that the entire intonation phrase is not split up by any boundaries. There should ideally be pauses or interviewer stretches before and after every IP.

- Make sure that the boundaries have been set correctly and as accurately as possible. You might have to zoom in a little bit more than usually in order to check on the accuracy.
- Listen to each intonation phrase in its whole. Check whether all words that are said in that part have actually been inserted in the corresponding interval. Delete all additional words or add all missing words.
- If you think that an intonation phrase should be taken out entirely because, for example, it is incomplete, and then you should mark that with the corresponding speech event or non-speech event abbreviation in your IP tier with your initials.

- **Intermediate phrases**
  - Listen to each ip and decide whether you would split it up as the first annotator did and whether all words that are transcribed in the interval are actually correct. Delete all additional words or add all missing words.
  - Make sure that the boundaries have been set correctly and as accurately as possible. You might have to zoom in a little bit more than usually in order to check on the accuracy.
  - Make sure that pauses that divide intermediate phrases are marked on the ip tier. If there are any kinds of pauses (up or b) that do not split up the ips, make sure that they are marked on the *Misc tier*.
- **Tone tier (HL)**
  - Do not change any tones directly. If you disagree with a pitch or phrase accent, then put the one that you would choose instead in your tier with your initials. If you think that there should not be any pitch or phrase accents, then set a point tier with "X" which stands for the deletion of this pitch or phrase accents.
  - All -?, *?, %?, ?-, ?*, ?% should be cleared up at this stage. Either decide for the tones' deletion with X, or a suggestion for a different solution.
- **Miscellaneous tier (Misc)**
  - Pay special attention to the annotations in the miscellaneous tier: Especially hesitation phenomena (hp) are often forgotten: by hesitation phenomena we mean every self-correction, repetition (not for emphasis), stuttering, and a lot of filled pauses followed or mixed with silent pauses.
  - Remember that there can be only one speech or non-speech event that can be annotated. Annotations like "b/int/hp", for example, are not allowed. You have to decide for the most salient one. For example, if the interviewer backchannels the interviewee while the interviewee breathes in, the "b" annotation would be more important than the interviewers backchanneling. If the interviewee however breathes in and the interviewer says something to which the interviewee reacts then "int" would be chosen as the more "important" annotation. Any additional Misc tier context relevant annotations should go in the Com Tier.
  - You do not need to listen to every stretch marked as a speech or non-speech event in any of the 4 tiers (IP, ip, Misc). You can take it as correct if it has been marked as a speech or non-speech event by the first annotator. However,

if you feel that a short intonation phrase is a response to a question and the question (int) has not been marked, then listen to the stretch before the IP again and relabel it if necessary.

- **Comments tier (Com):** If you write a comment put your initials behind the comment.
- **Feedback for annotators:** If you notice more or less systematic errors in the annotation that you are checking, then please inform the annotators and the whole team (in the CC) on the issue via E-Mail or in our next team meeting.

Step 14. A2 - "final" annotation check

- Check whether the tiers have been named correctly and are in the right order (IP, ip, HL, Misc, Com)
- Important: Do not delete any tiers by anybody
- Look at the IPs first and listen to them. Decide whether you would go for option 1 (A0 annotator) or option 2 (A1 annotator). If you agree with option 2 (A1 annotator) then incorporate the changes in the upper tiers by the A0 annotator.
- Then repeat the same procedure for the tones
- If you would not agree with any of the two options add a comment to the Com Tier and bring the file to our next team meeting and discuss the critical points with the other annotators.
- When you are still unsure about segmentation of phrases and tones then please leave a comment in the Com Tier.
- If you decide for either one of the options by A0 or A1 then adjust the miscellaneous tier if necessary and the tone tier, because if intermediate boundaries change then a tone might have to changed from H- to H* for example.
- Check whether the annotation symbols have been deleted (only "=" (for incomplete words), "<X>" "<XX>" "<XXX>" (for one or more unclear words, each X standing for a word), "<foreign>…<foreign>" (for foreign words), and "[i:]"are allowed) everything else should be deleted. Especially ":" and "<?>" are mostly overseen.
- Check whether all *erm, em, ehm, uhm* have been replaced with *um*, and all *eh* with *er*, and all *mhm* with *mm*.
- Delete all double spaces in the text file as well as the *TextGrid* file.
- Check whether all comments have been taken care of but leave the Com Tier in
- Find out the final word count by copying all IPs into a word document after you have decided on one of the options for all IPs.

Text file

- Each text file should have exactly the same words as they appear in the IPs in the *TextGrid*.
- In order to get the correct words, you should copy each phrase from the intonation phrase tier into a word document. That way we can be sure that it's a 100% the same and it is easier than correcting the old text file.
- Make sure the files are saved as RTF files.

Step 15. One time-events: Inter-annotator agreement and intra-annotator agreement

- Agreement between all annotators annotating the same recording (inter-annotator agreement), and the agreement for each annotator with himself or herself annotating the same recording twice (intra-annotator agreement) will be conducted.
- Both procedures will take place once the annotators have gained more experience with annotation. Approximately after 6 months - 1 year of annotation the annotators will annotate a recording that they have annotated at the beginning. Additionally, they will annotate a recording that has been previously annotated by another annotator (Karin Puga) after at least 6 months of annotation experience.
- Every annotator involved in the project annotated CZ32 for inter-annotator agreement. Additionally, everybody chose one of the files that were annotated at the beginning (one of the first files) and once again after 1 year.

Appendix 2. Figures

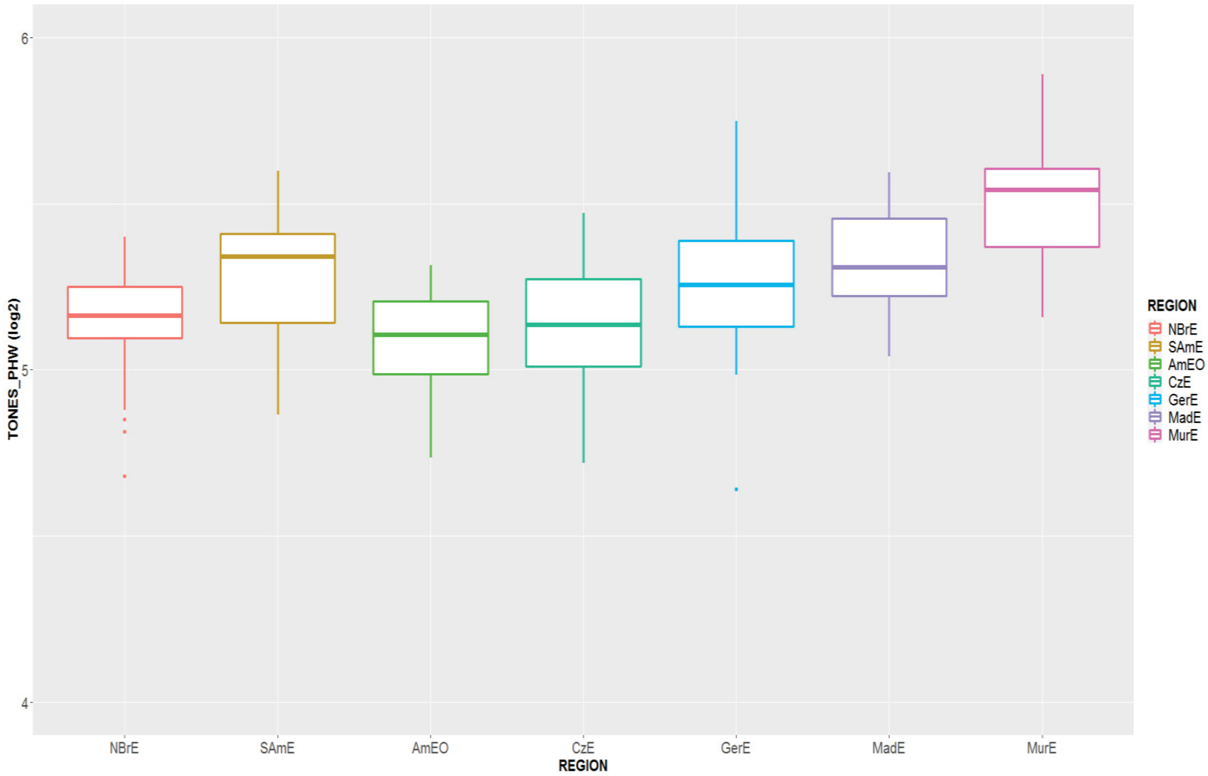


Figure 54. Boxplot for number of tones (phw) by REGION

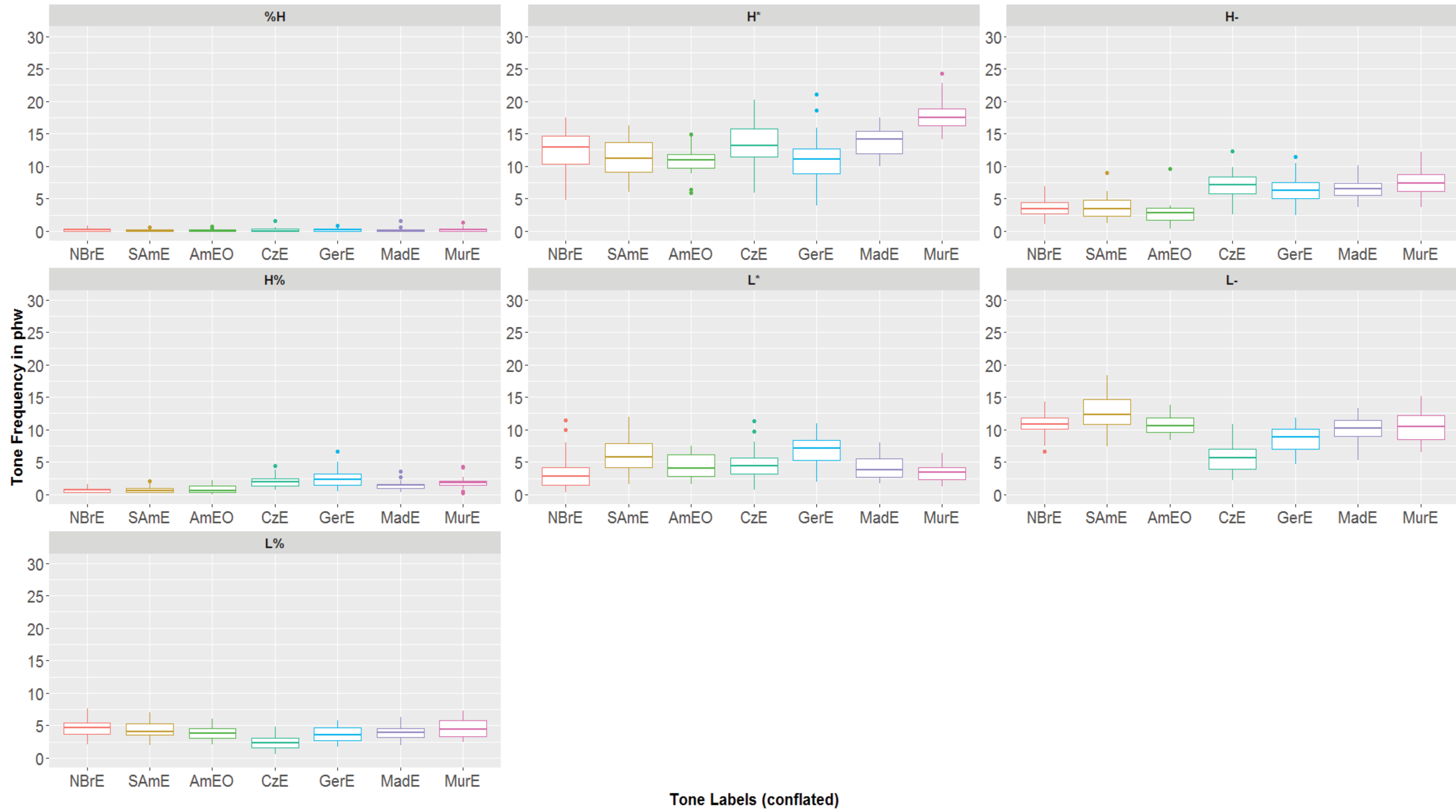


Figure 55. Boxplots for frequency of tone labels (normalized in per hundred words) by tone height (High (H) or Low (L)) by position within IPs and ips (“*”=pitch accent, “-“=phrase accent, “%”=boundary tone) by REGION. Tones are conflated, i.e. H* includes: !H*, H*, ^H*, L+H*, L* includes: L*, H+L*, L*+H, H- includes: !H-, H-, ^H-, and H% includes: !H%, H%, and ^H%, %H, L- and L% do not include any other levels

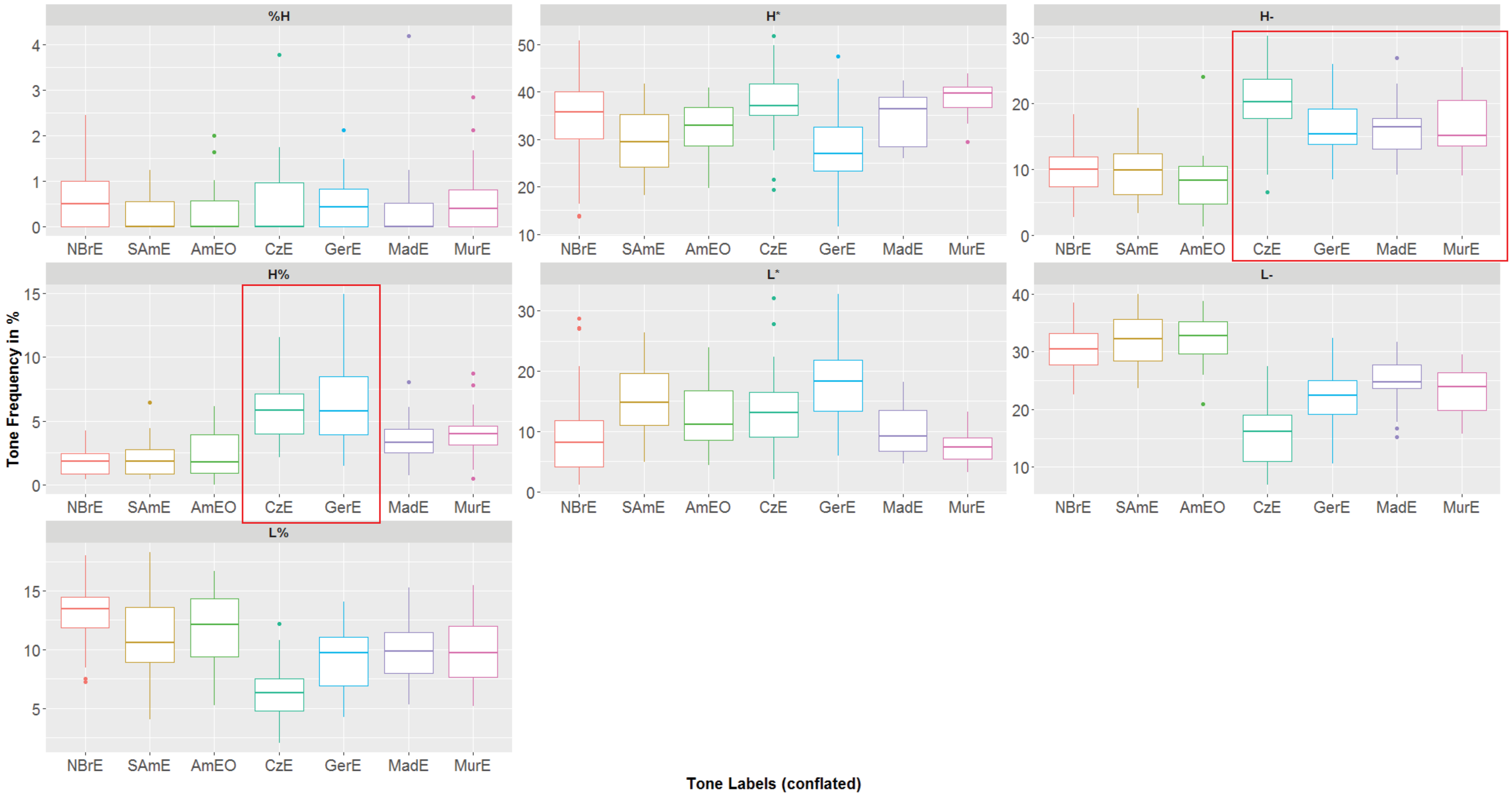


Figure 56. Boxplots for frequency of tone labels (in percent) by tone height (High (H) or Low (L)) by position within IPs and ips ("*"=pitch accent, "-"=phrase accent, "%"=boundary tone) by REGION. Tones are conflated, i.e. H* includes: !H*, H*, ^H*, L+H*, L* includes: L*, H+L*, L*+H, H- includes: !H-, H-, ^H-, and H% includes: !H%, H%, and ^H%, %H, L- and L% do not include any other levels

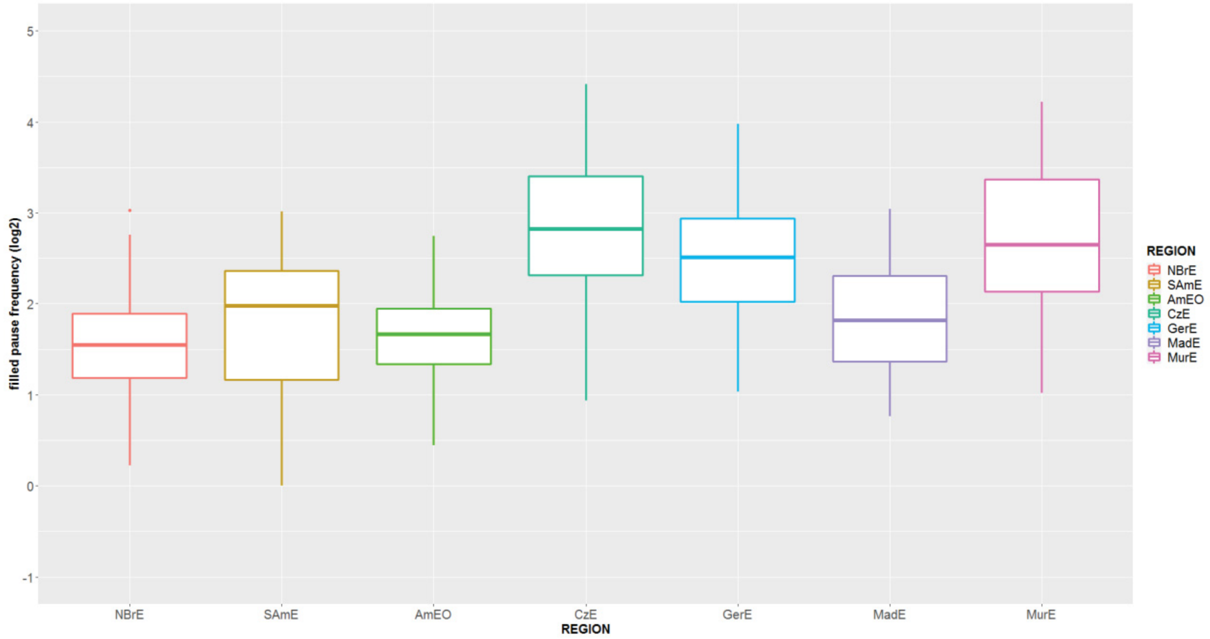


Figure 57. Boxplot for number of all filled pauses phw (logged) (including *er*, *um*, and *mm*) by REGION

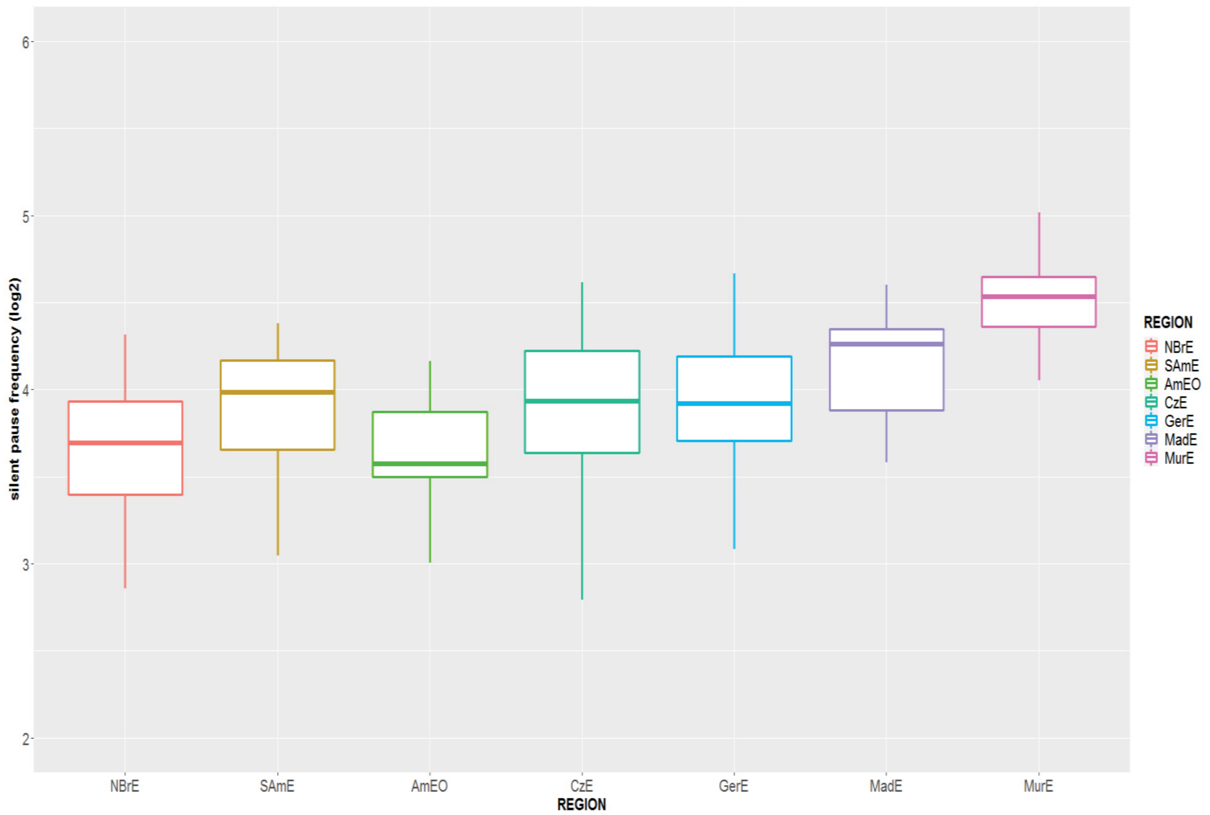


Figure 58. Boxplot for number of all silent pauses phw (logged) (including *up* and *b*) by REGION

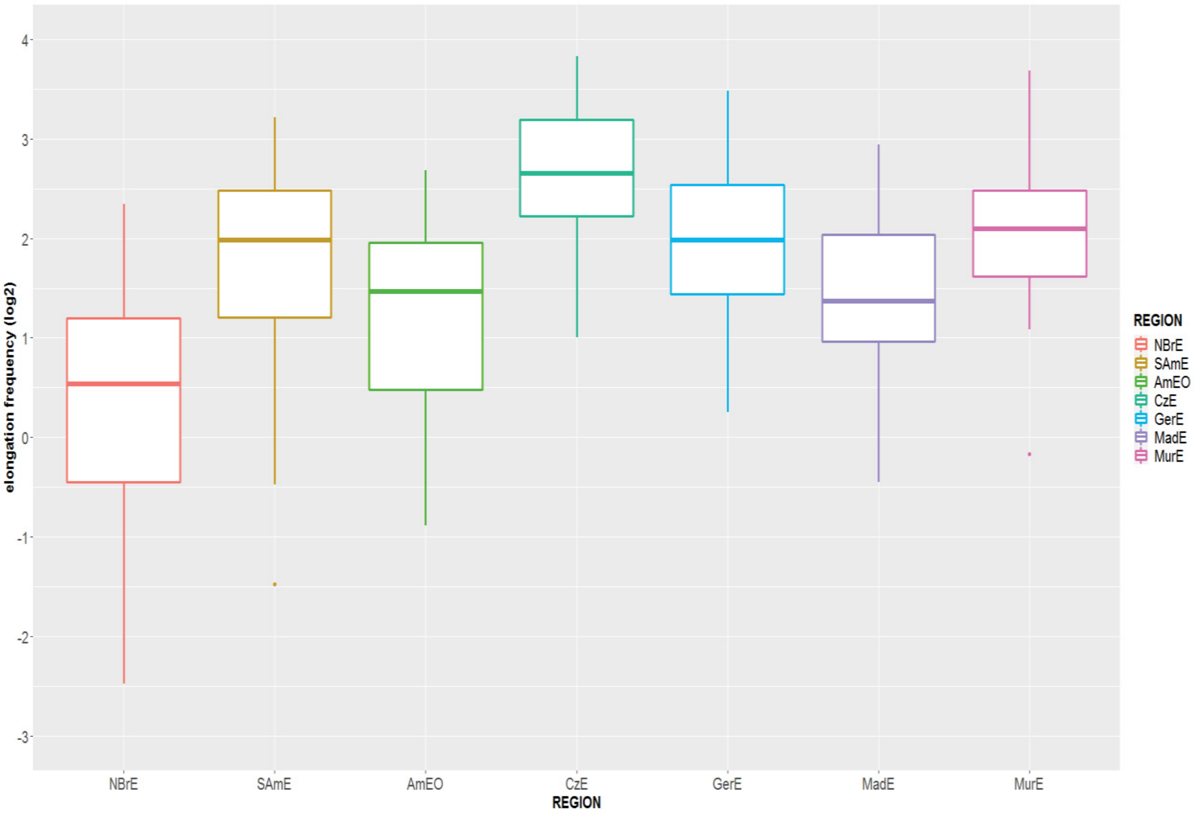


Figure 59. Boxplot for number of elongations phw (logged) by REGION

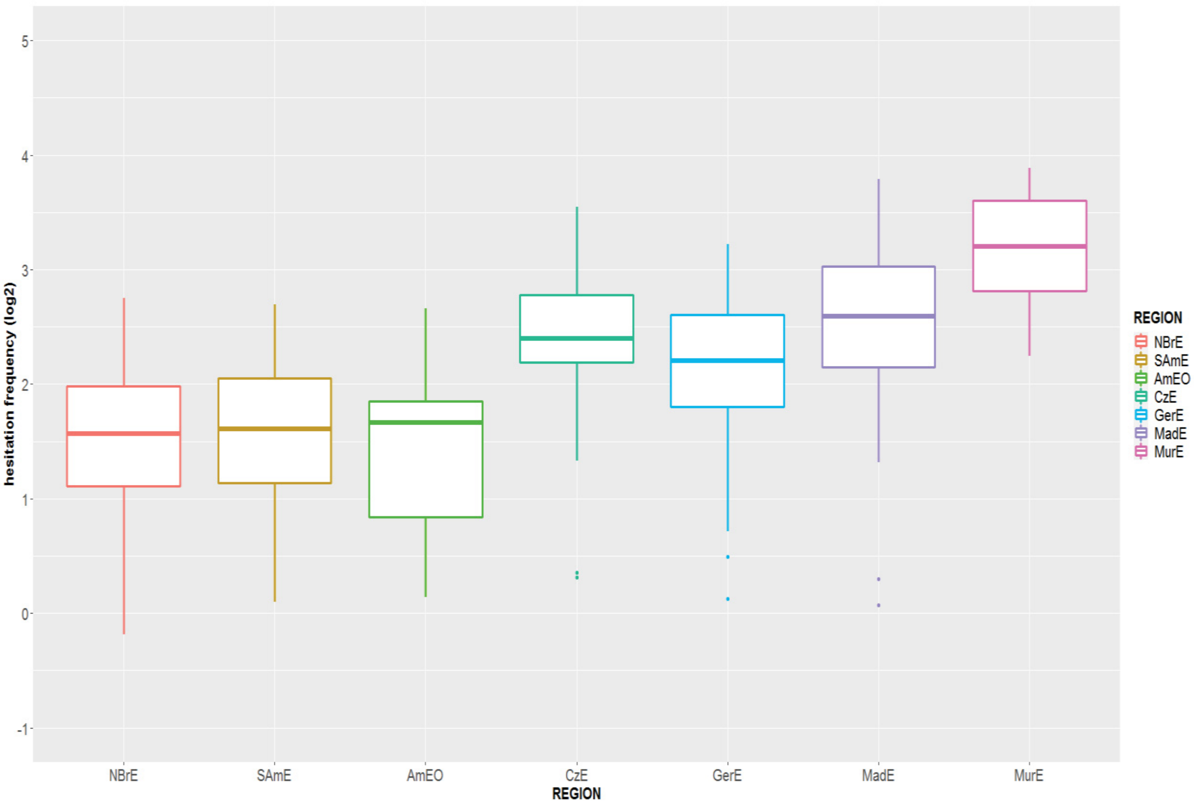


Figure 60. Boxplot for number of hesitations phw (logged) by REGION

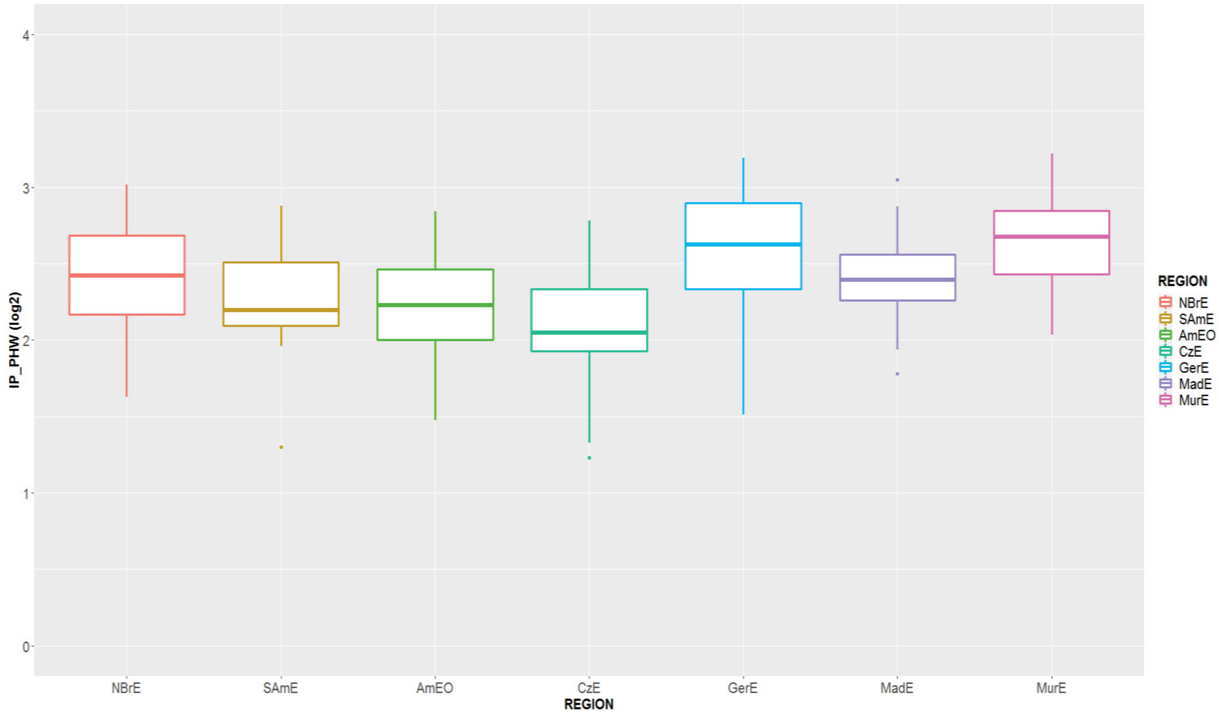


Figure 61. Boxplot for intonation phrases (IPs) by REGION, normalized in phw (number of IP by speaker divided by the number of words per speaker *100) with logged frequencies of IP phw log2(IP/WORDS*100), Spanish regional dialect (Madrid n=25 speakers, Murcia n=20 speakers). American regional dialect (South n=29, Mid-Atlantic=6, North n=4, New England n=2, Midland n=2, West n=2)

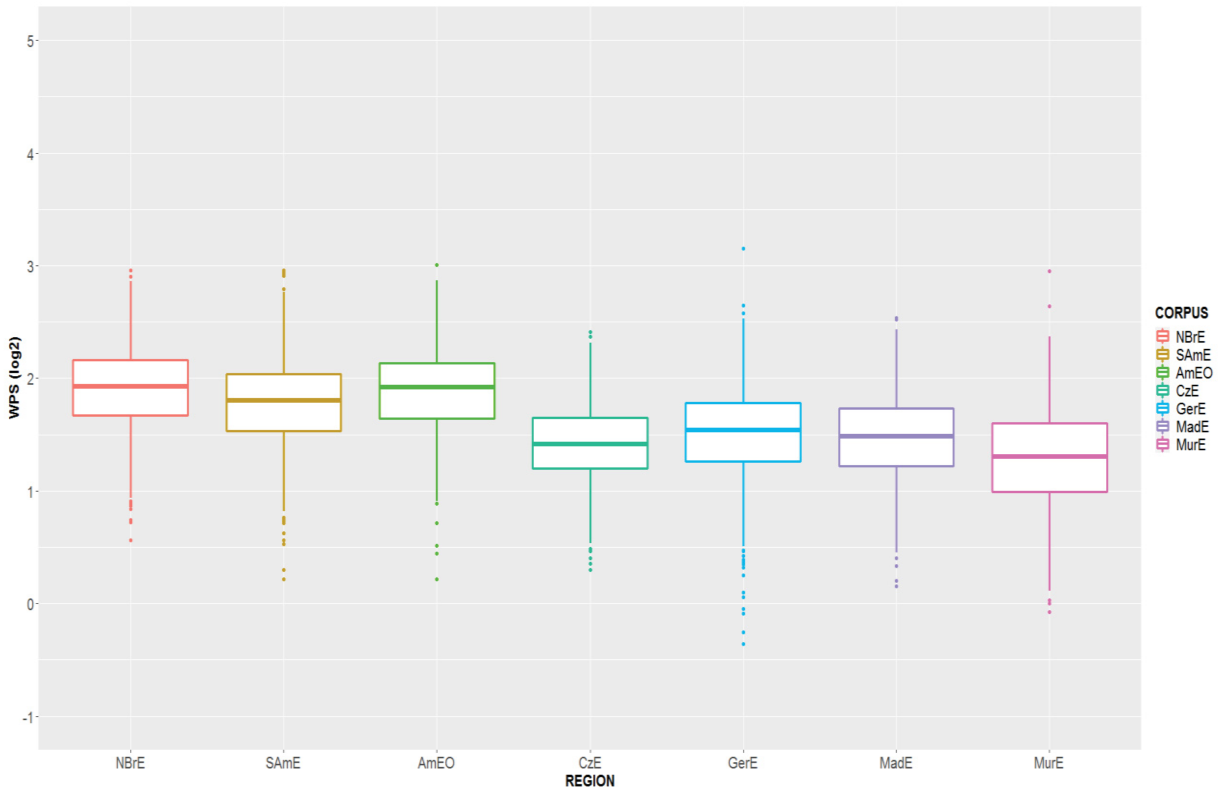


Figure 62. Boxplot for IP length in WPS (logged values). Normalized: number of words of each IP/seconds of each respective IP

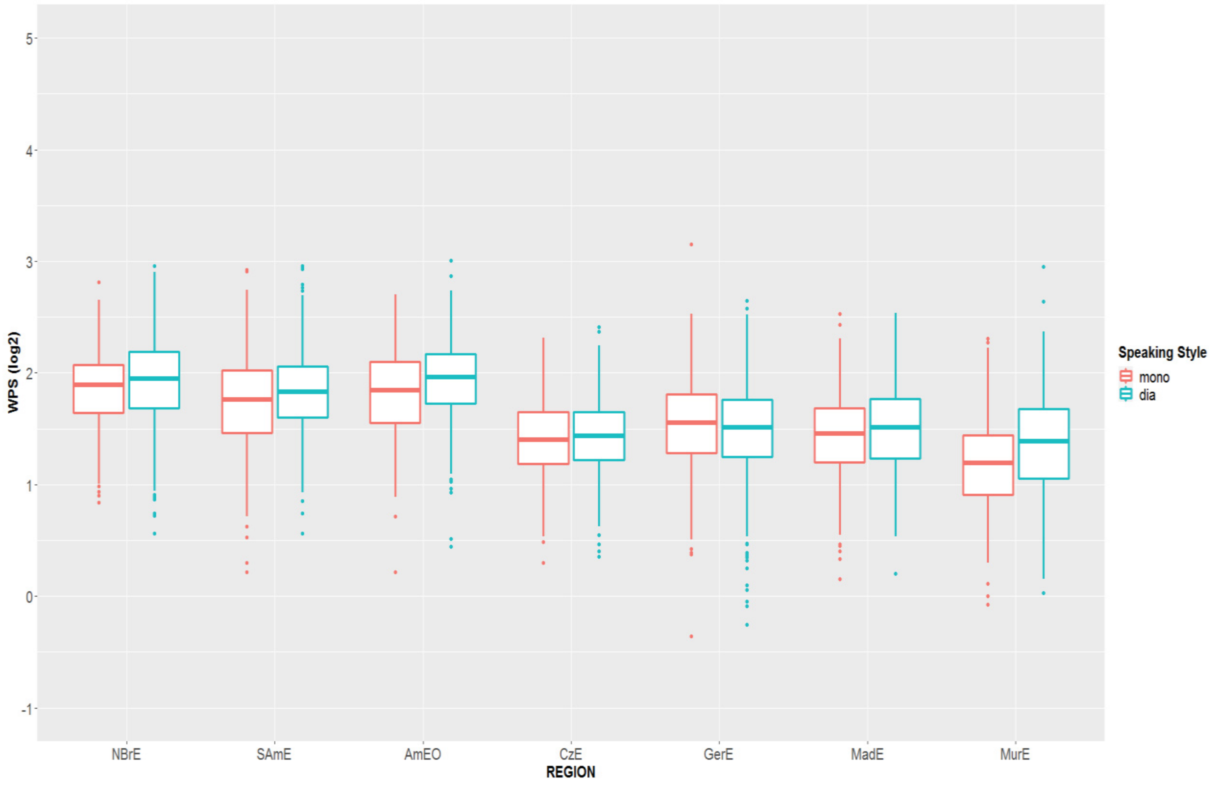


Figure 63. Boxplot for IP length in WPS (logged) in the two speaking styles (monologues vs. dialogues)

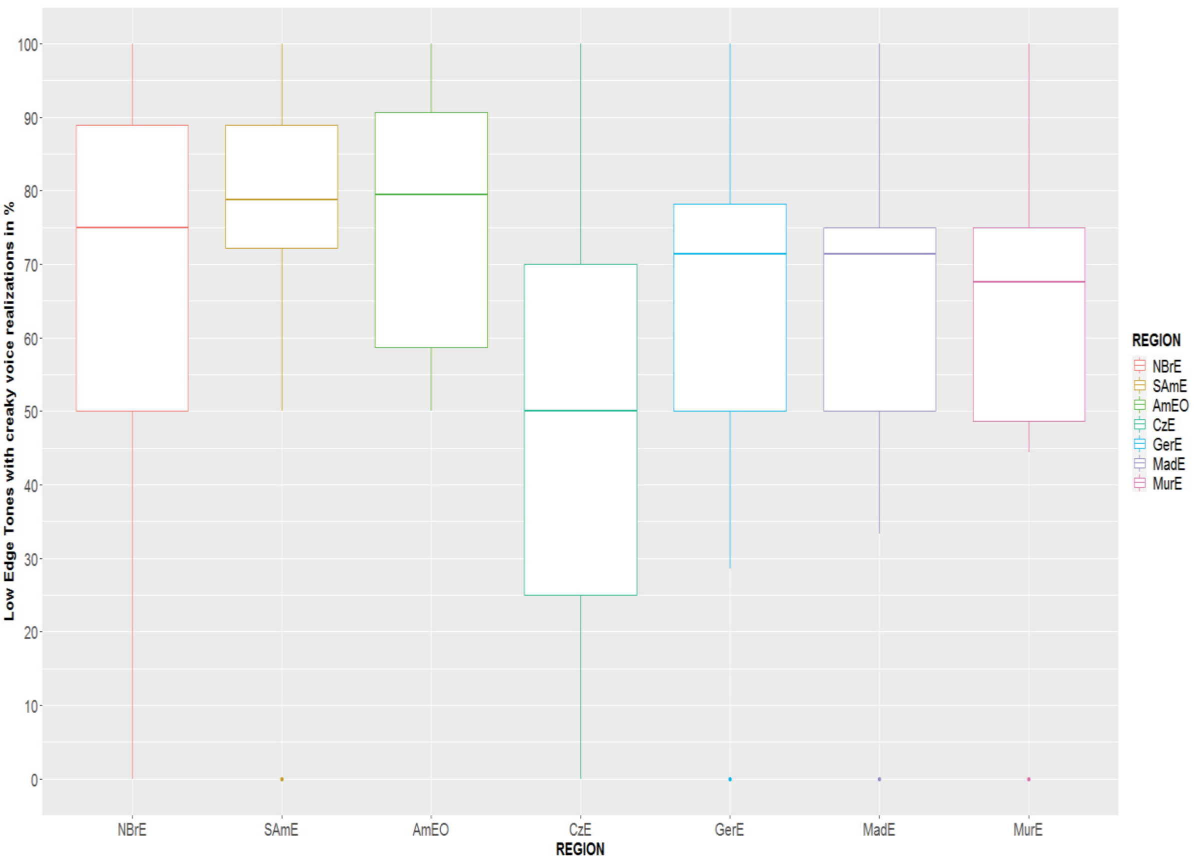


Figure 64. Boxplot for percentage of low edge tones produced with a creaky voice by REGION

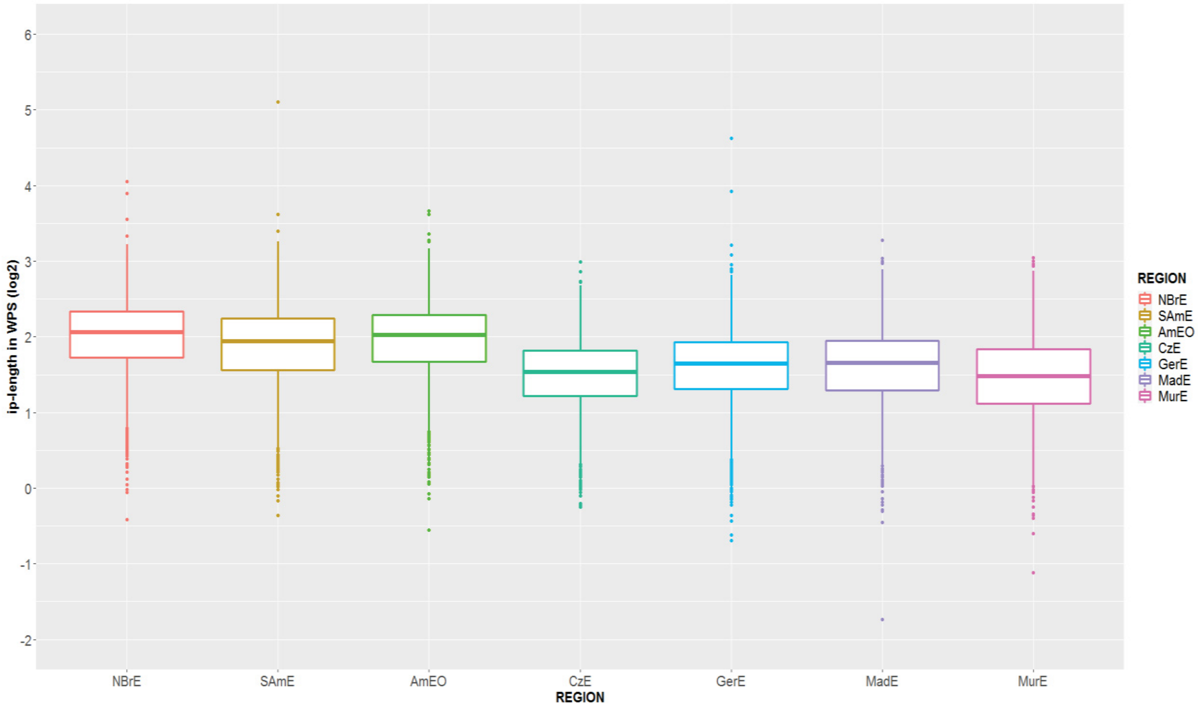


Figure 65. Boxplot for ip speech rate in words per seconds (WPS)

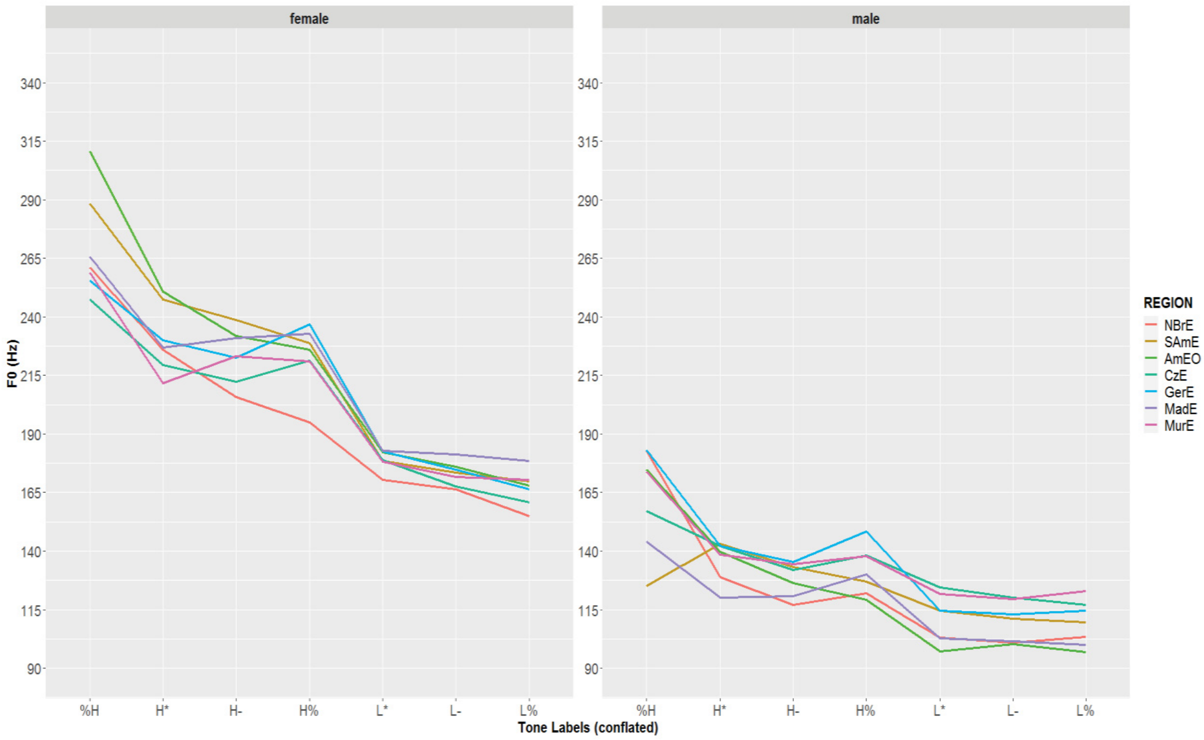


Figure 66. Line plots for median f0 height by TONE LABEL (conflated), REGION, and SEX, with speaker CZ008 included in the female CzE group

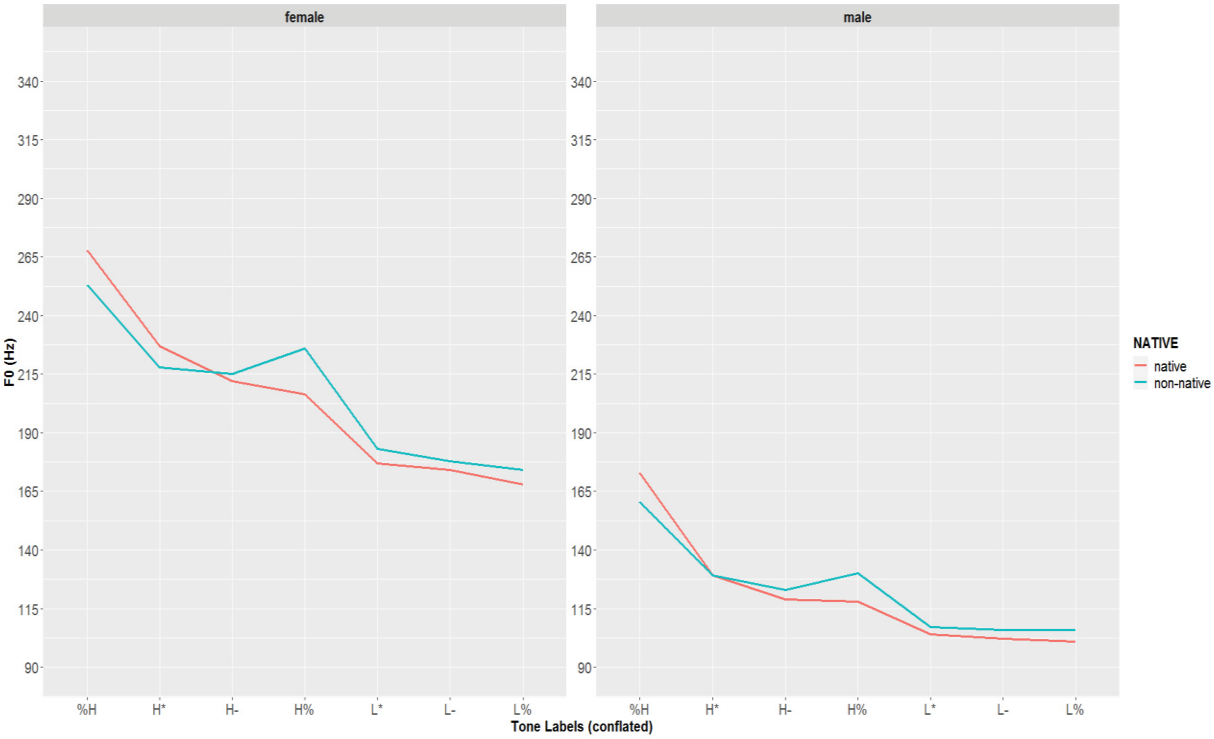


Figure 67. Line plots for median f0 height by TONE LABEL (conflated), speaker status and SEX, with speaker CZ008 included in the female CzE group

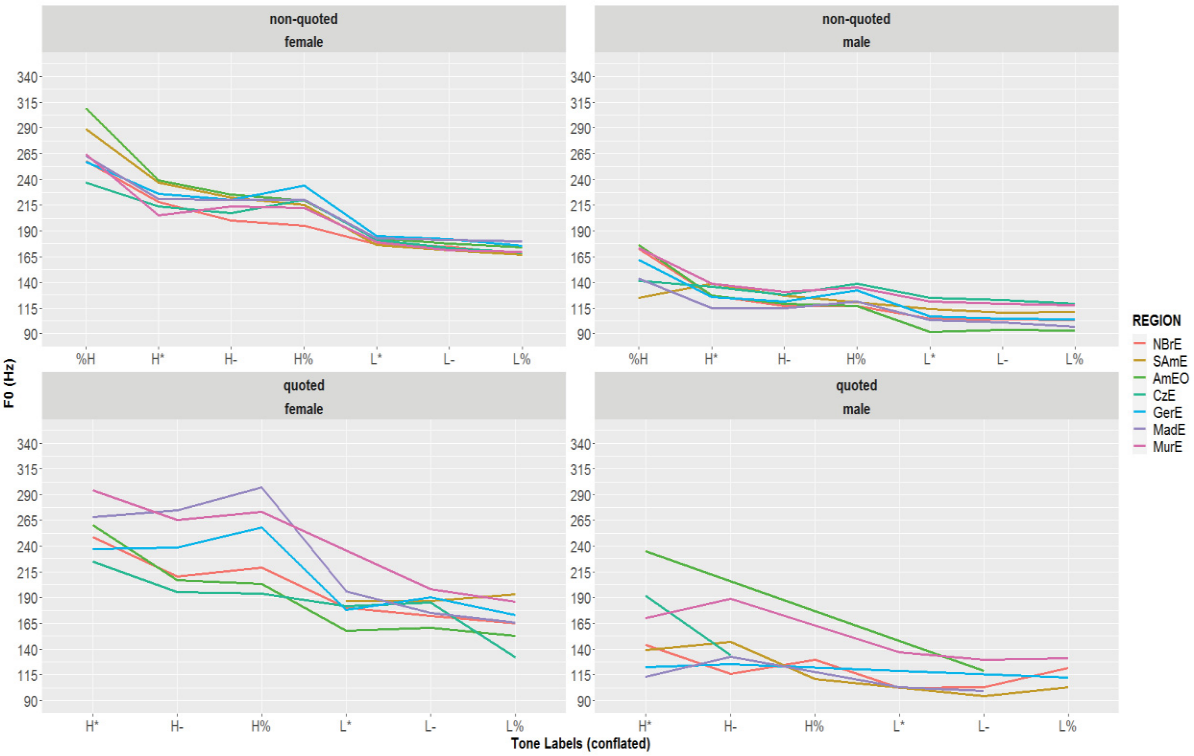


Figure 68. Line plots for median f0 height in quoted vs. non-quoted speech by REGION and SEX

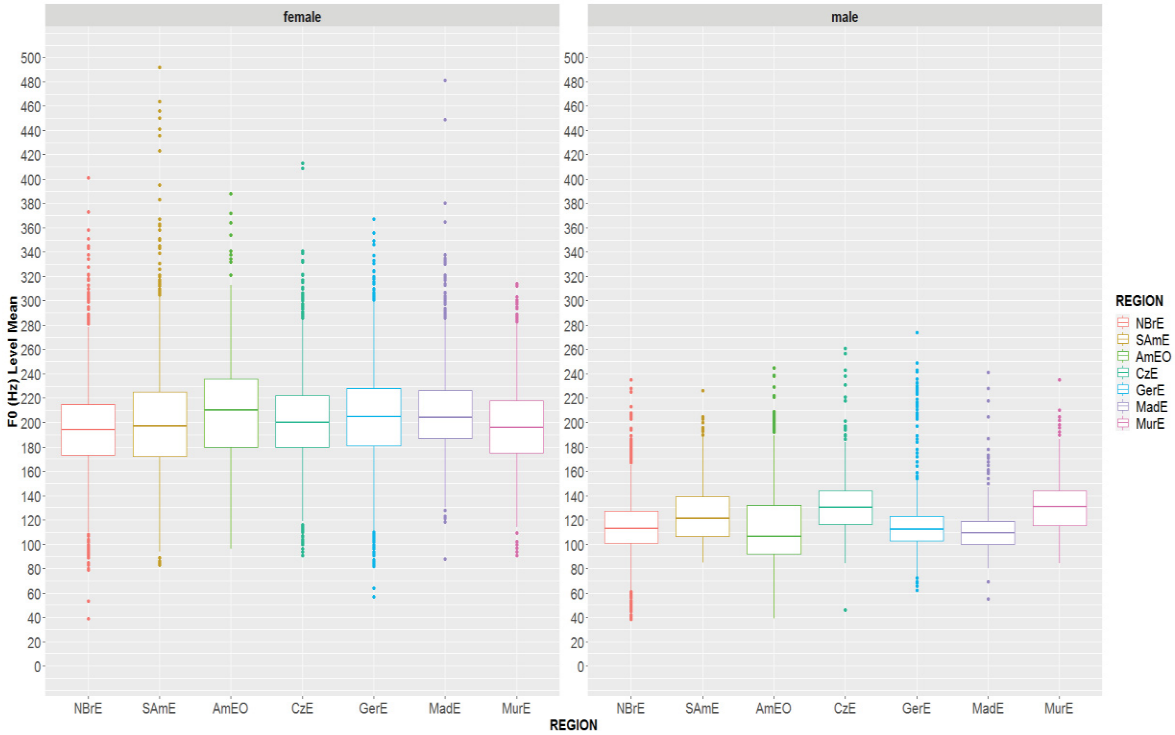


Figure 69. Boxplots for f0 level (mean) by REGION and SEX

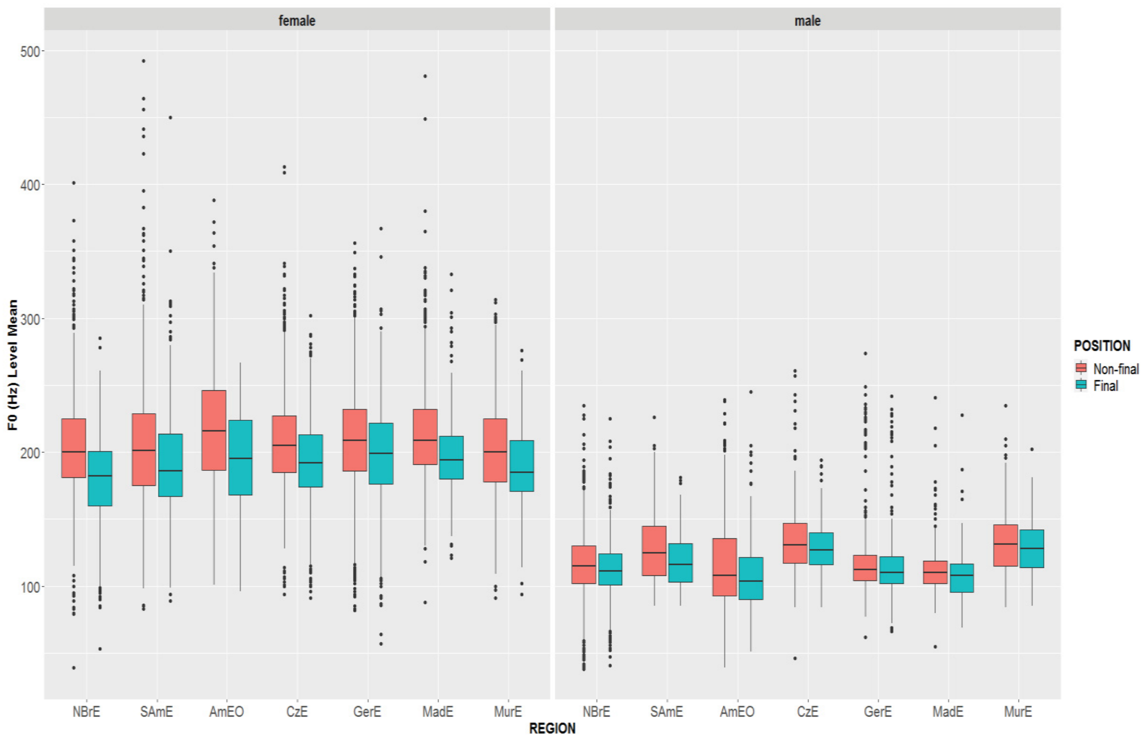


Figure 70. Boxplots for f0 level (mean) by REGION, SEX, and POSITION (IP internal vs. IP-final)

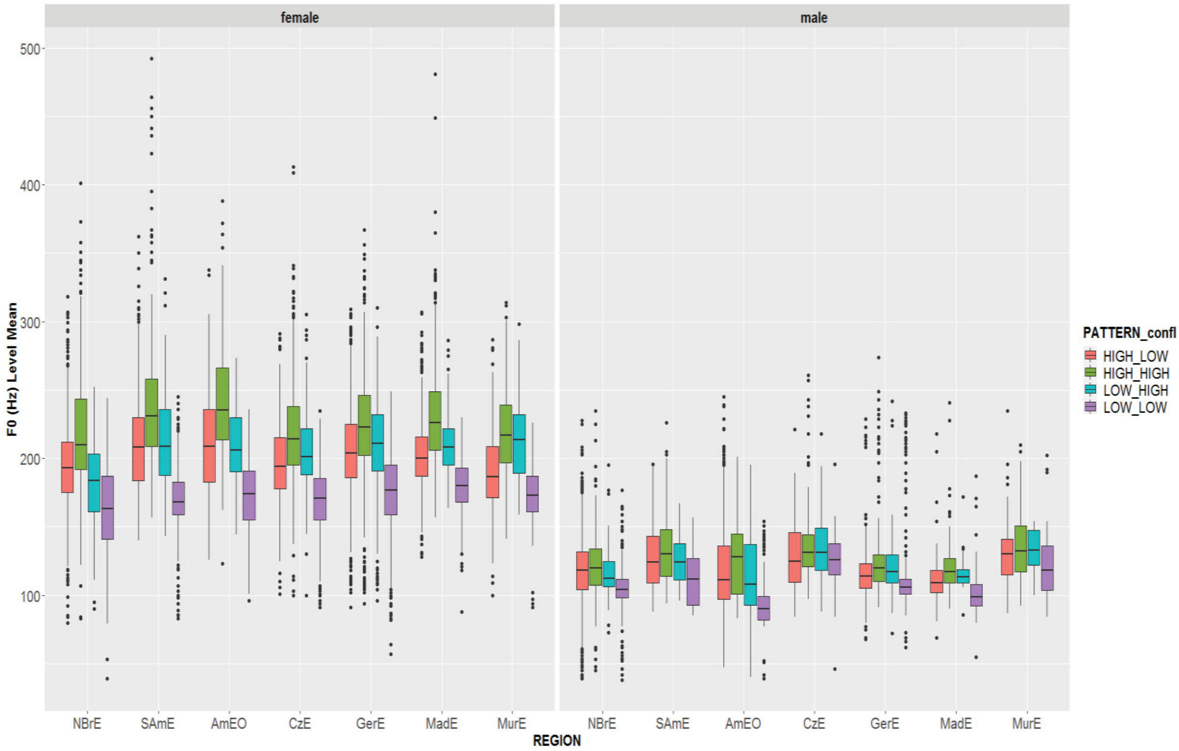


Figure 71. Boxplots for f0 level (mean) by conflated tunes (TUNE_PATTERN), REGION, and SEX

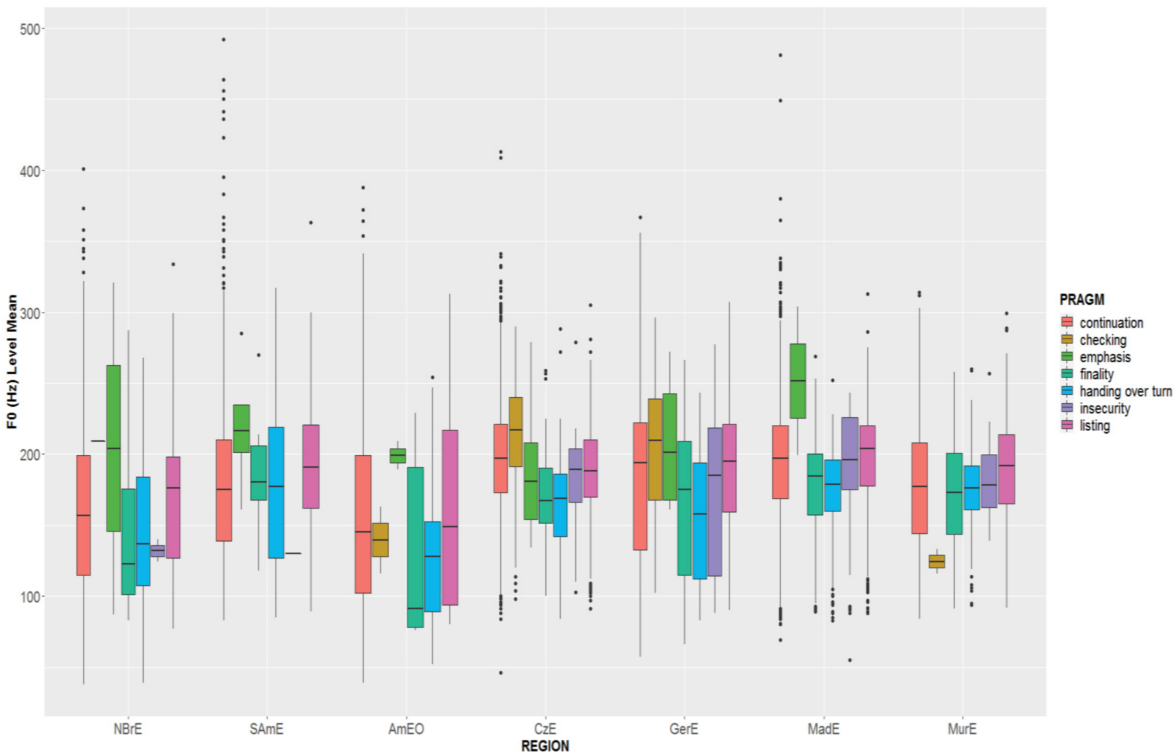


Figure 72. Boxplot for f0 level (mean) by REGION and PRAGM

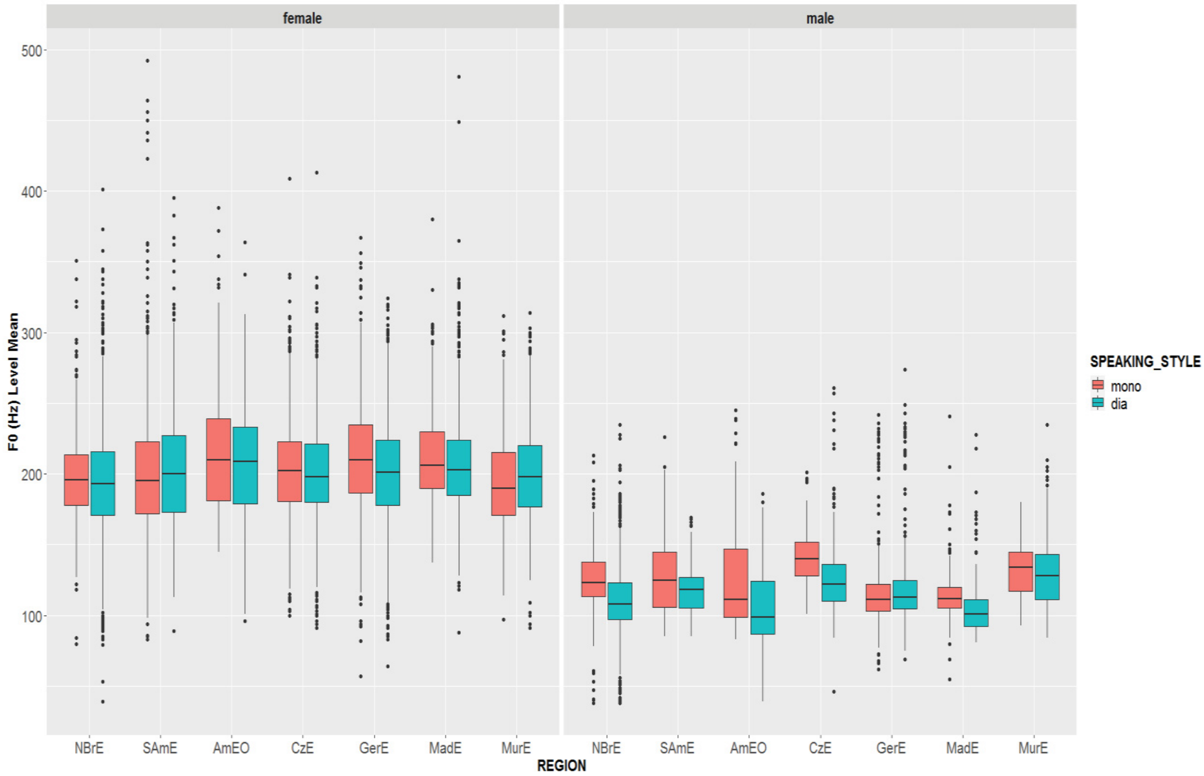


Figure 73. Boxplots for f0 level (mean) by REGION, SEX, and SPEAKING_STYLE

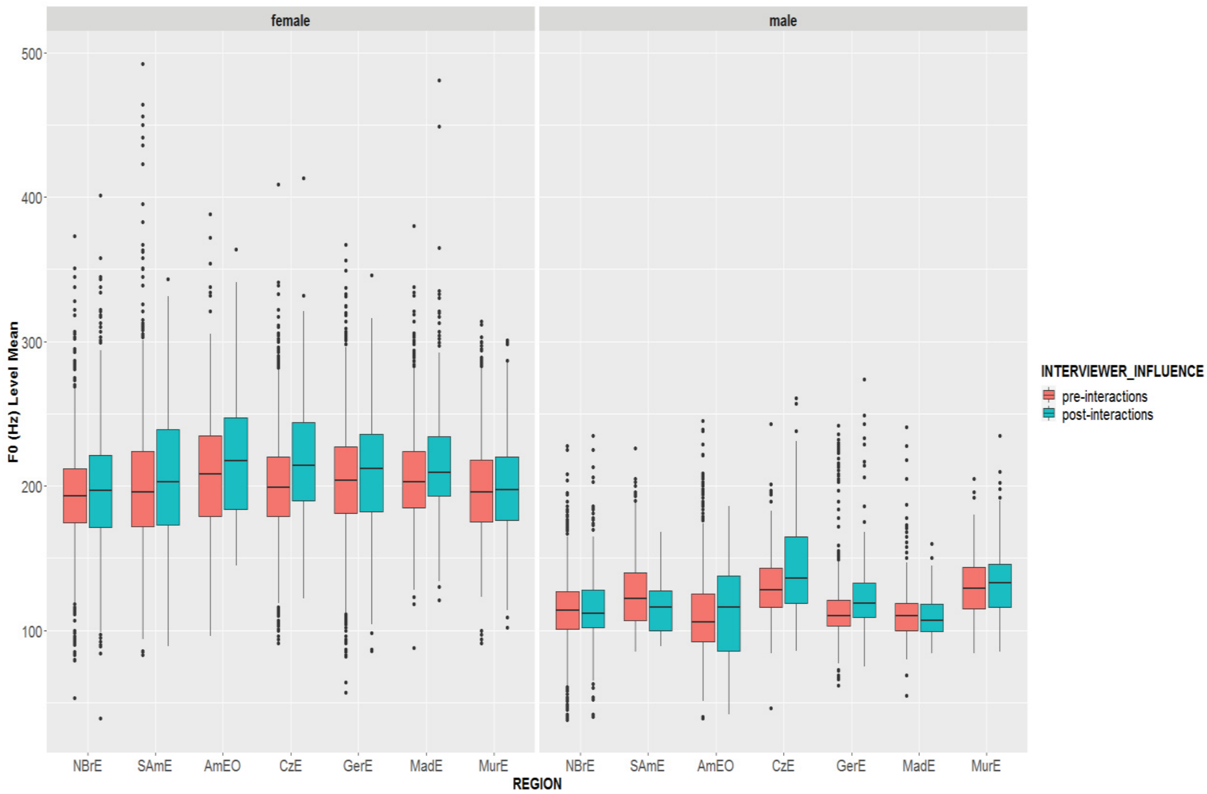


Figure 74. Boxplots for f0 level mean by REGION, SEX, and INTERVIEWER_INFLUENCE

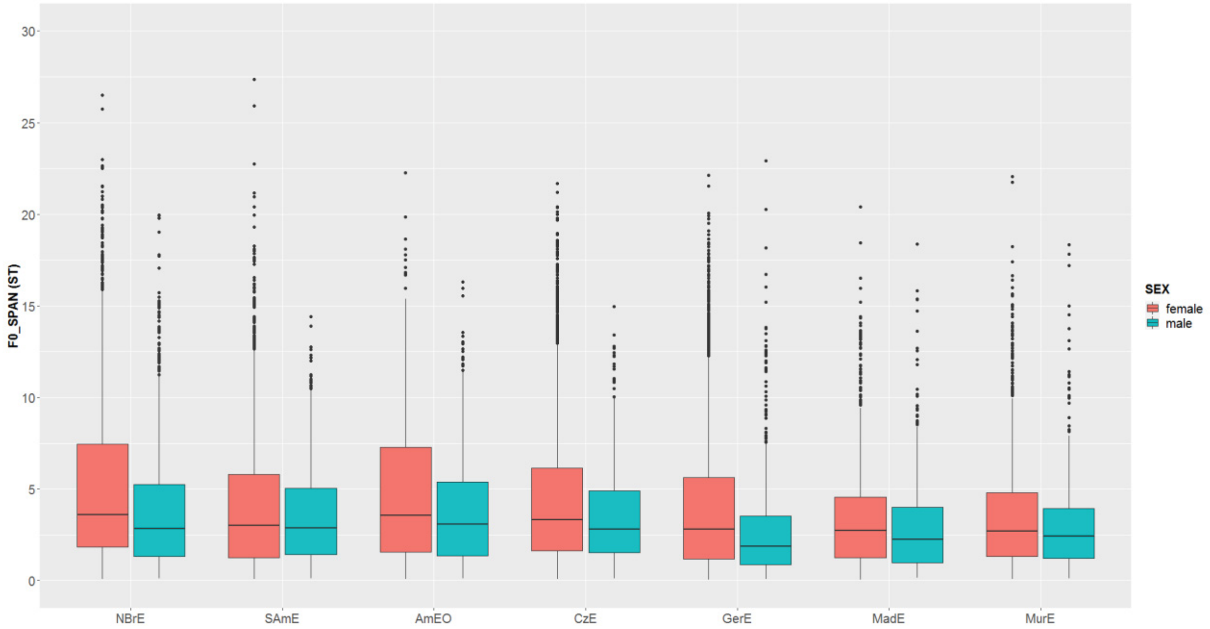


Figure 75. Boxplot for f0 span in STs by REGION and SEX. Semitone conversion was calculated with the following formula: 12*log2(F0_MAX/ F0_MIN)

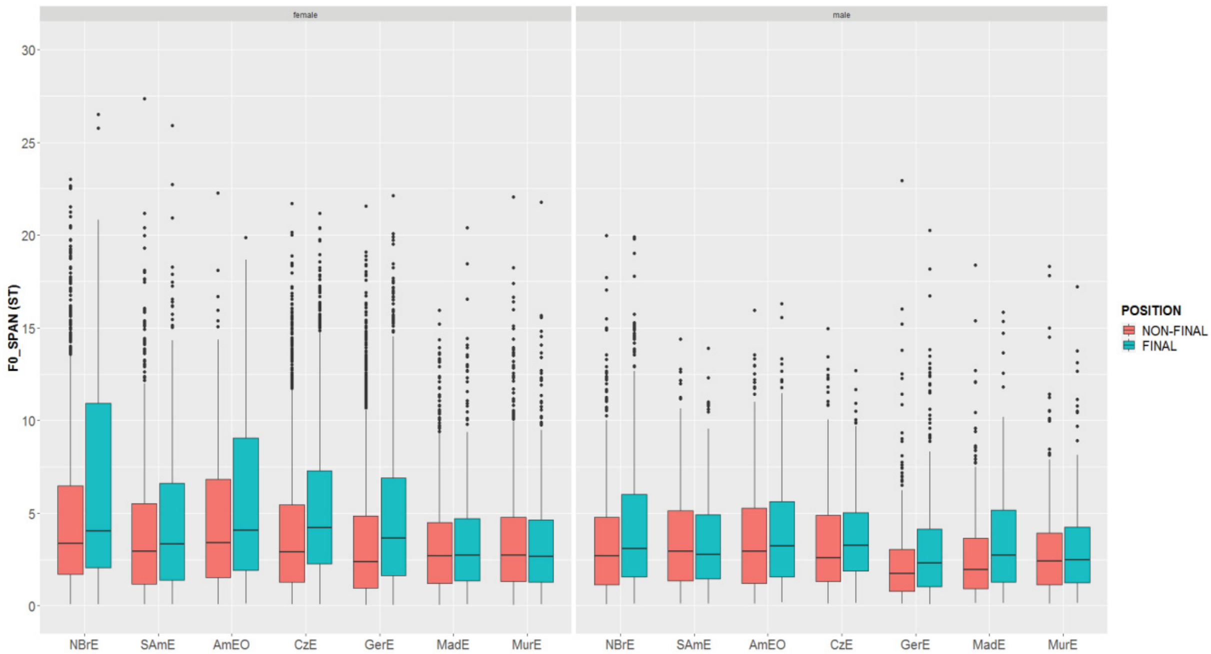


Figure 76. Boxplots for f0 span (in STs) by REGION, SEX, and POSITION (IP internal vs. IP-final)

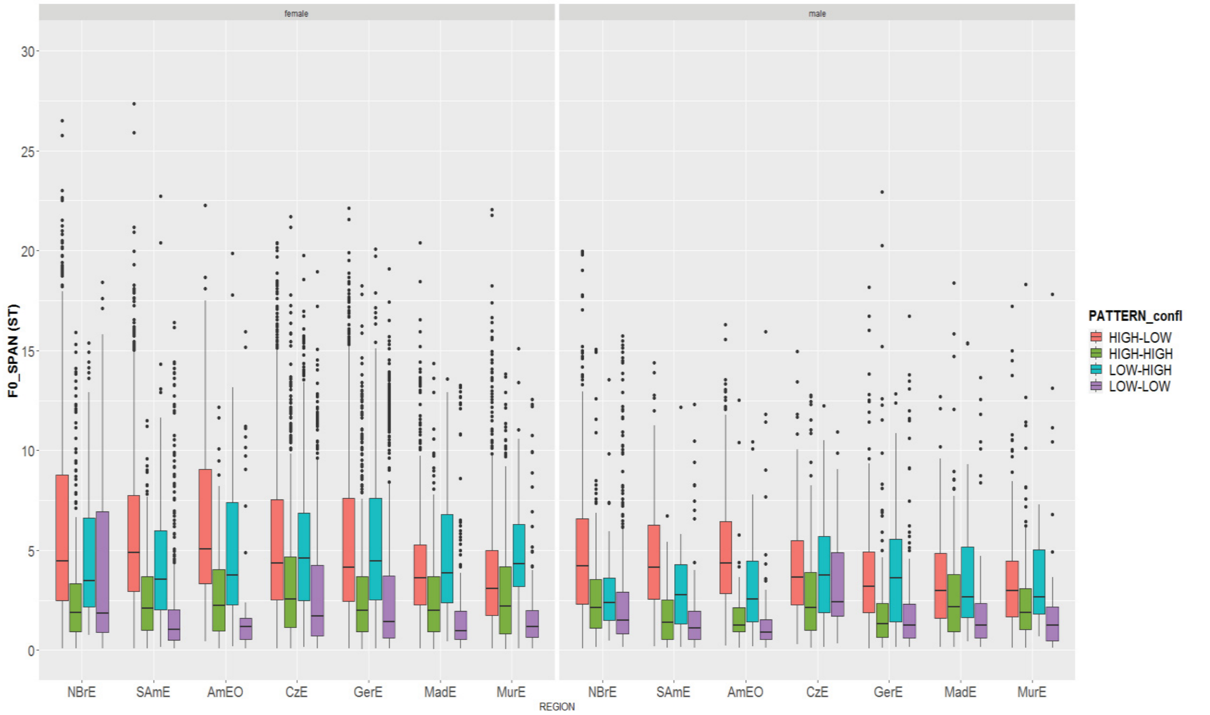


Figure 77. Boxplots for f0 span (in STs) by conflated tunes (PATTERN_confl), REGION, and SEX

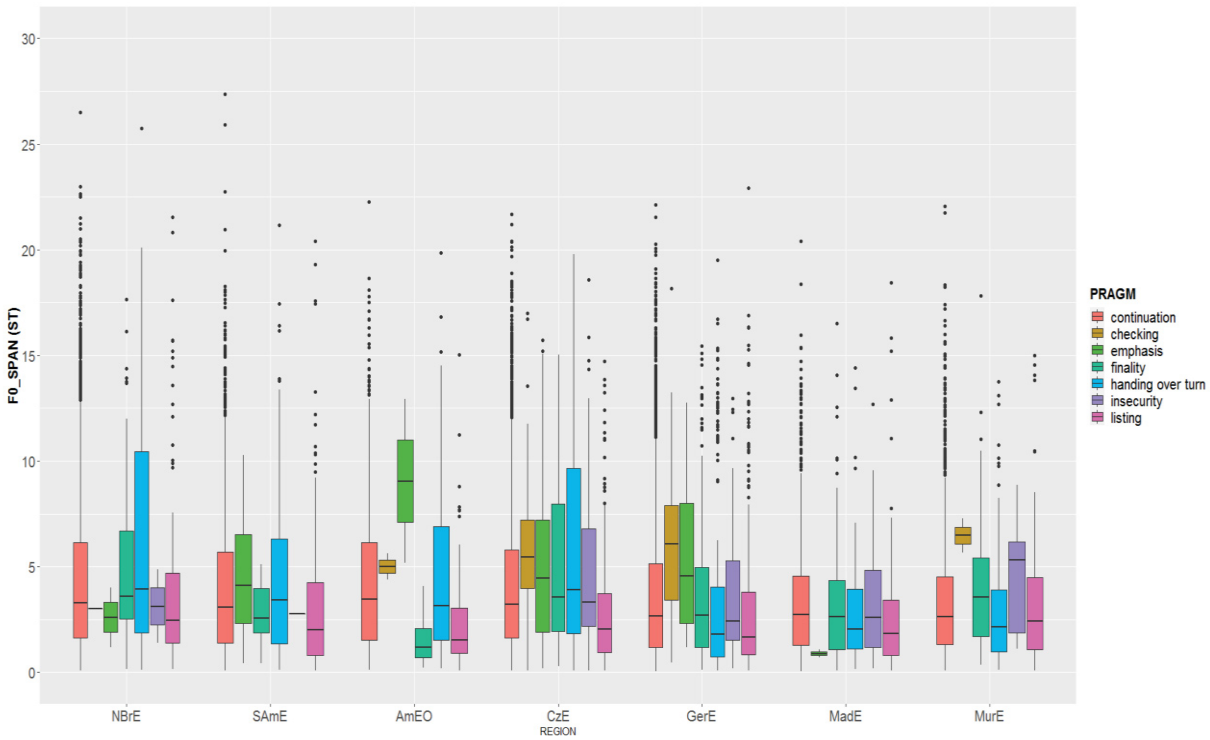


Figure 78. Boxplot for f0 span (in STs) by REGION and PRAGM

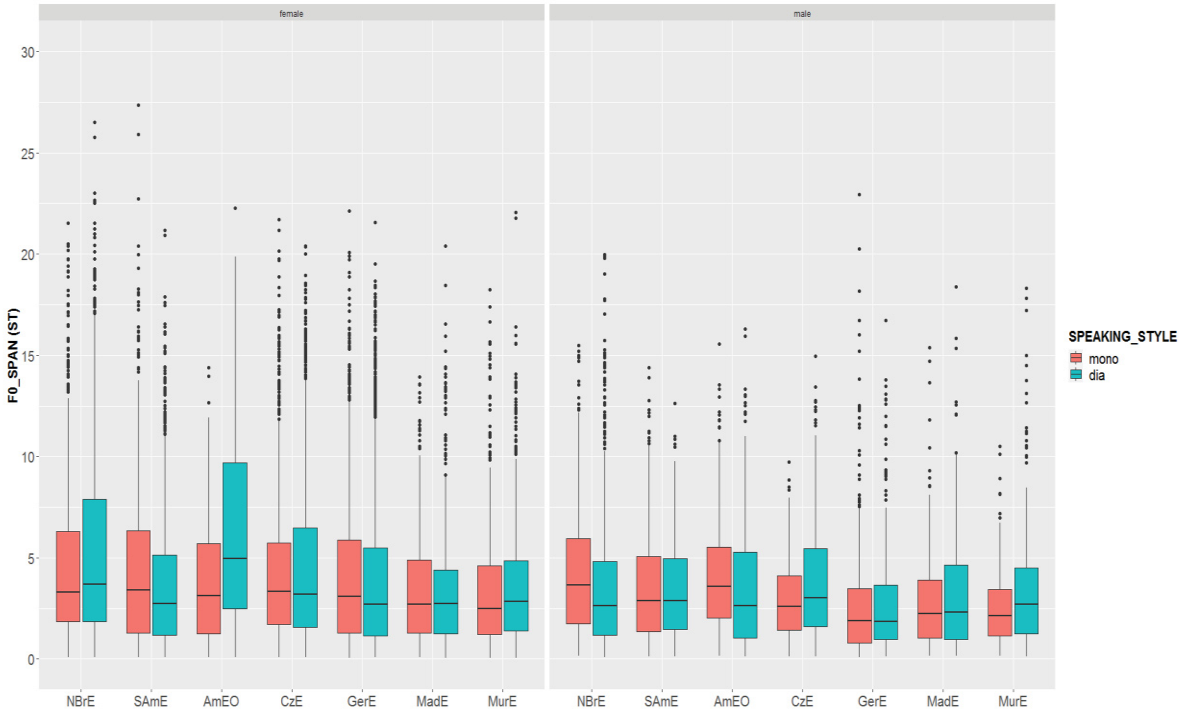


Figure 79. Boxplots for f0 span (in STs) by REGION, SEX, and SPEAKING_STYLE

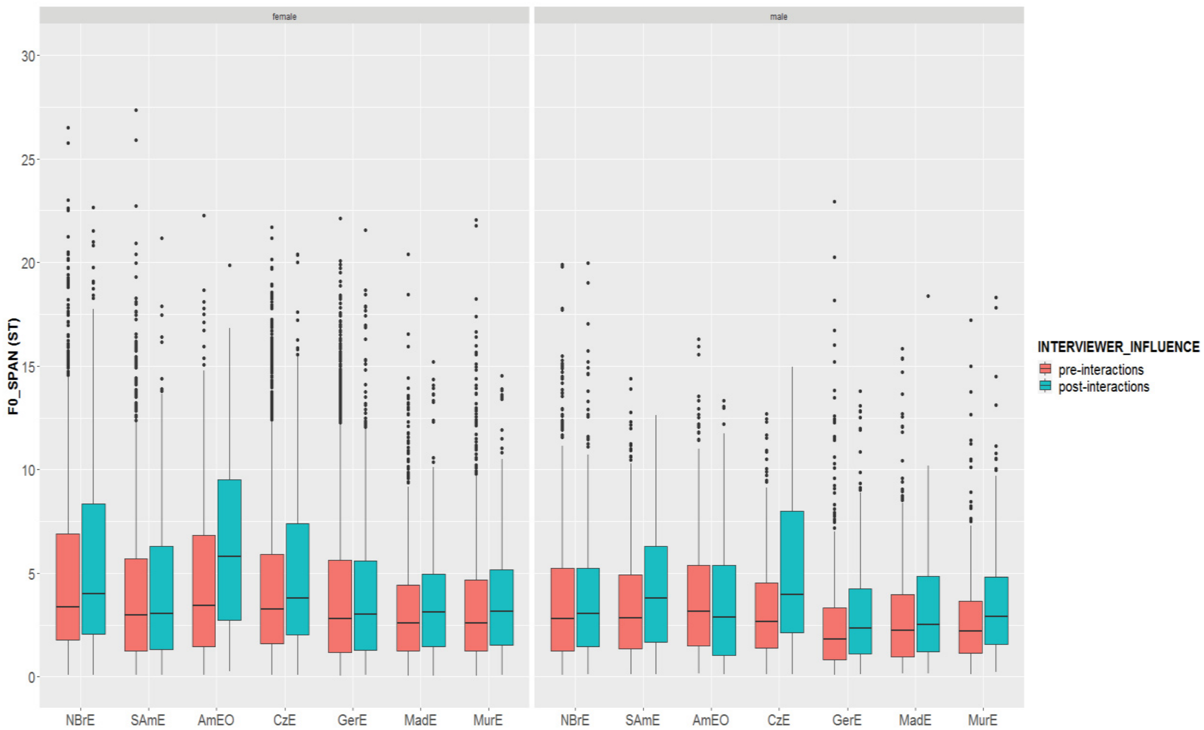


Figure 80. Boxplots for f0 span (in STs) by SEX, REGION, and INTERVIEWER_INFLUENCE

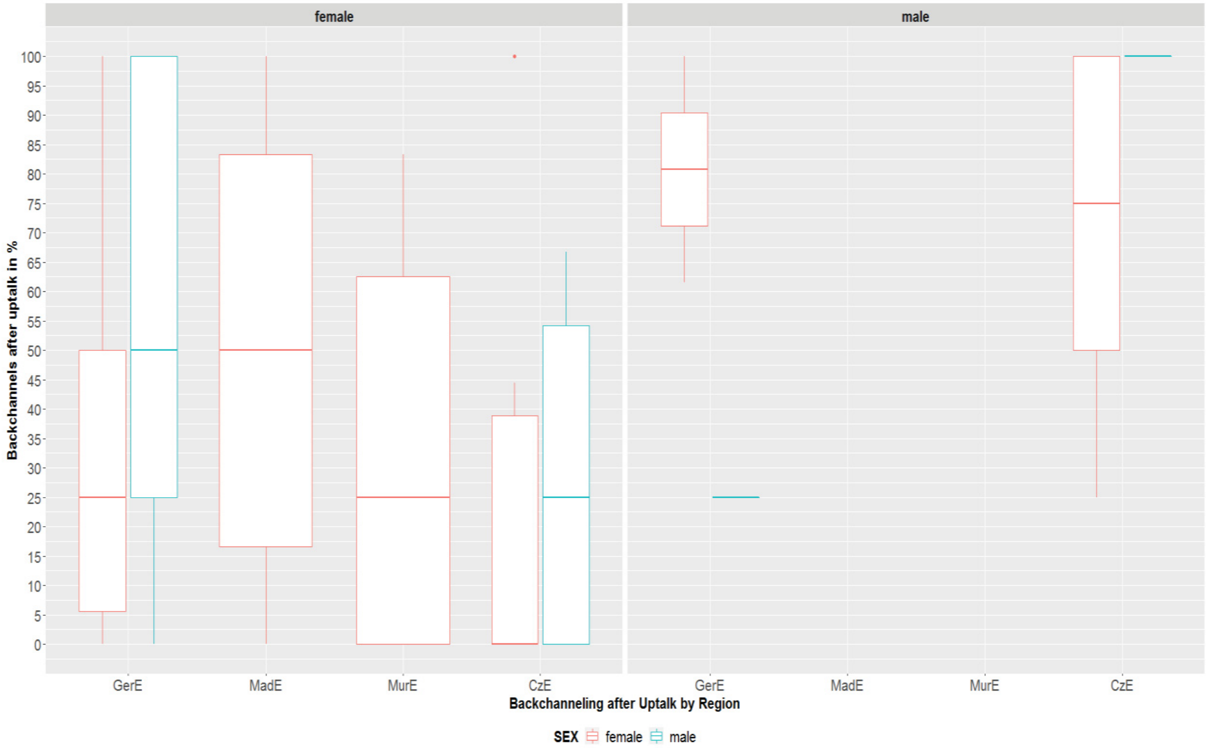


Figure 81. Boxplots for audible backchanneling after uptalk instances by REGION and SEX

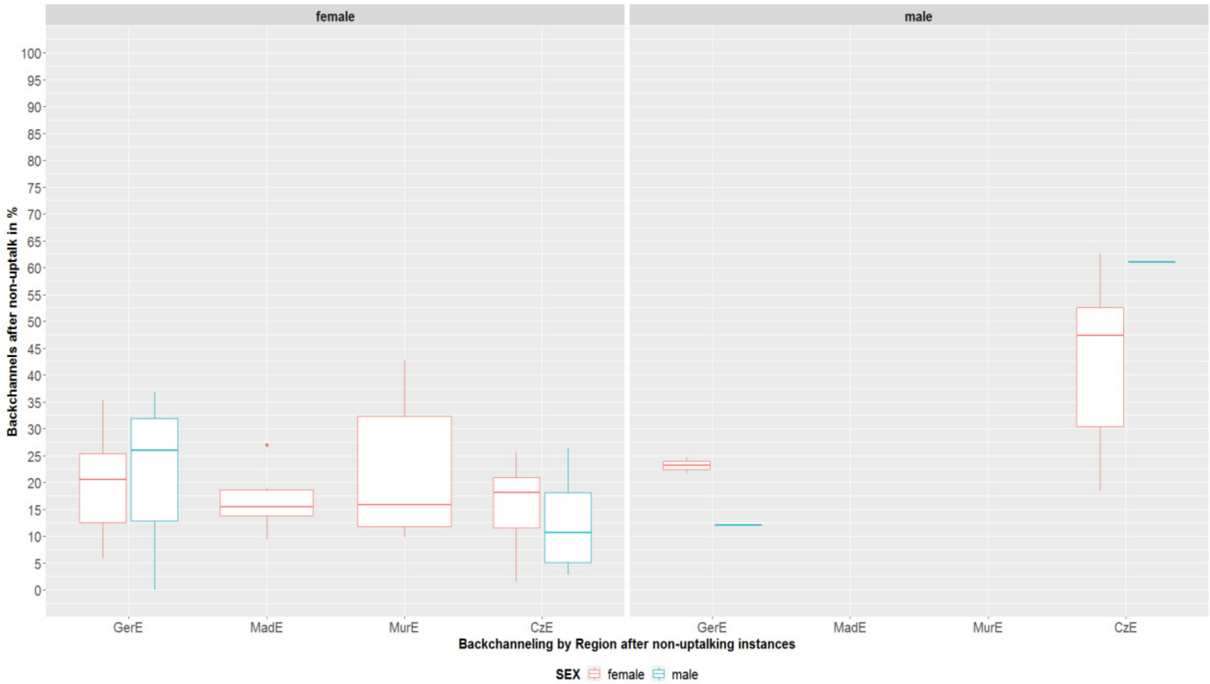


Figure 82. Boxplots for audible backchanneling after non-uptalking ips by REGION and SEX

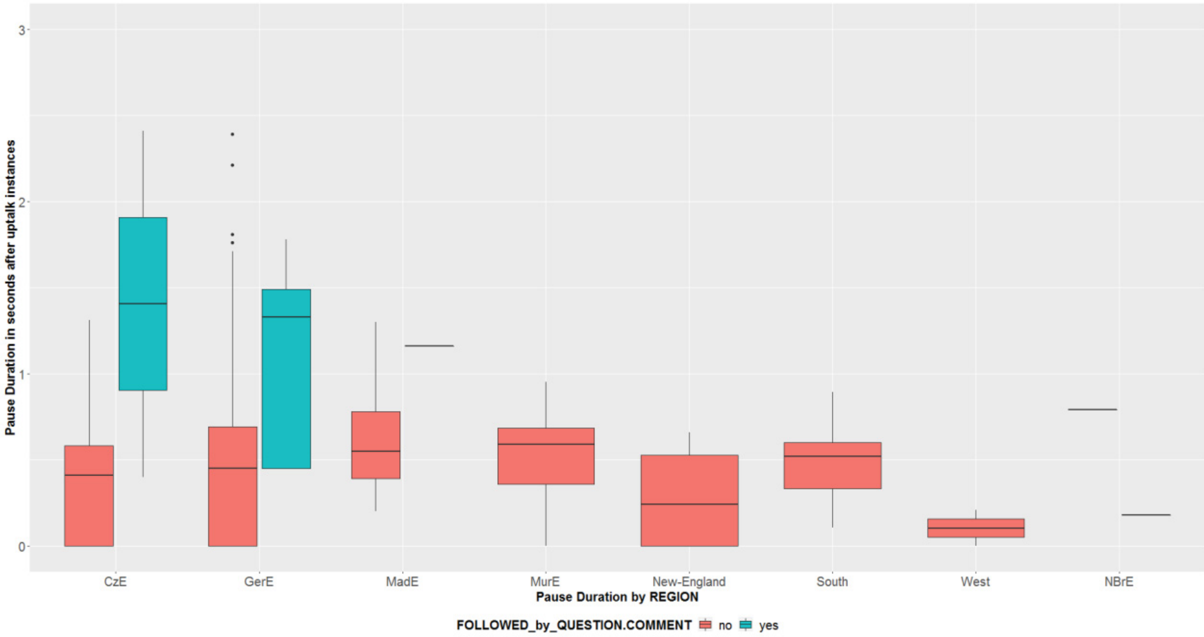


Figure 83. Boxplot for pause duration in seconds after uptalk instances by REGION and interviewer interventions

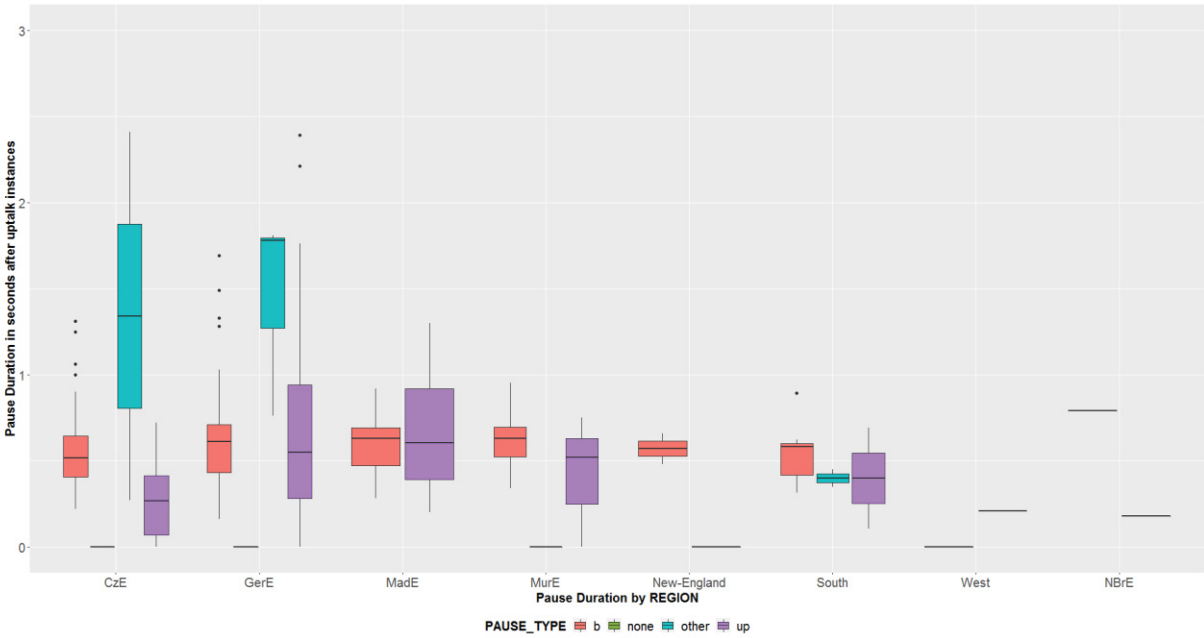


Figure 84. Boxplot for pause duration in seconds by REGION and pause type

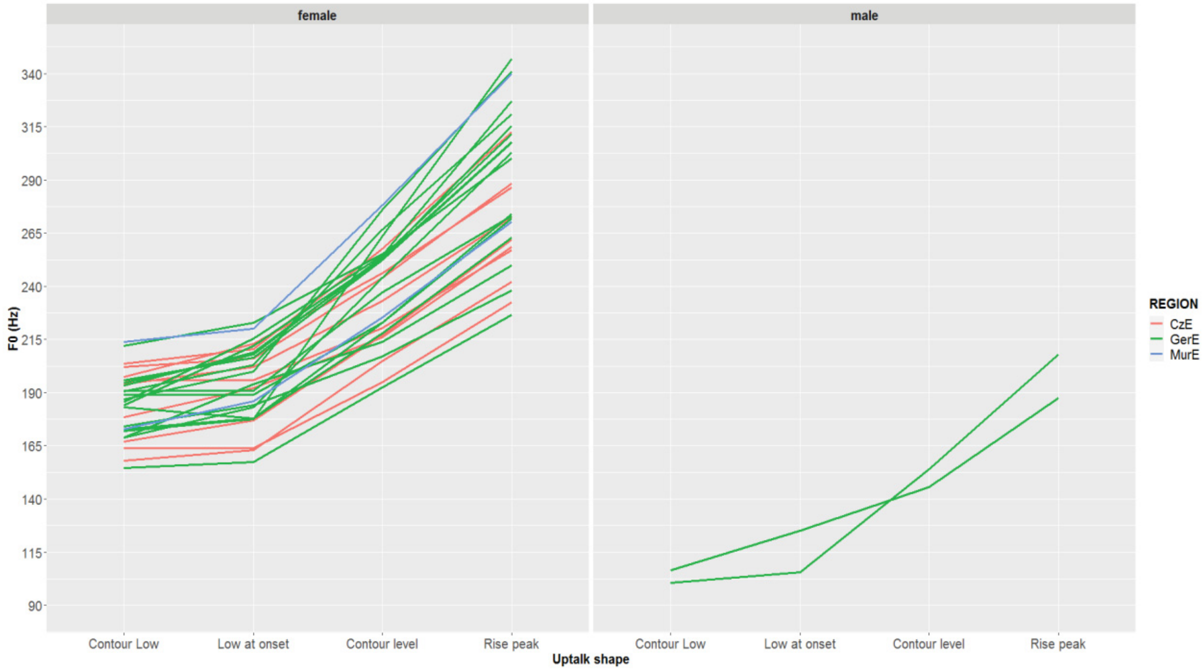


Figure 85. Line plots for f0 tune lows, low at onset, tune level, rise peak by REGION and SEX

Appendix 3. Tables

Table 54. Results for binomial logistic regression for H- and L-tones. Final model: HL ~ REGION * POSITION + PRAGM + INTERVIEWER_INFLUENCE + TOPIC + SEX * INT_SEX + UPTALK + QUOTED. Goodness-of-fit statistic (C=0.76). Baseline 1: 0.52, Baseline 2: 0.50. Accuracy=0.71, precision=0.71, recall=0.73

Fixed effects:	**Estimate**	**Simple odds ratios**	**Std. Error**	**p-value**	
(Intercept)	1.62	5.04	0.05	< 0.001	***
REGION_SAmE	-0.76	0.47	0.06	< 0.001	***
REGION_AmEO	-0.42	0.65	0.07	< 0.001	***
REGION_CzE	-0.27	0.76	0.06	< 0.001	***
REGION_GerE	-0.97	0.38	0.05	< 0.001	***
REGION_MadE	-0.22	0.80	0.06	< 0.001	***
REGION_MurE	0.24	1.27	0.07	0.0010	**
POSITION_edge	-2.68	0.07	0.05	< 0.001	***
PRAGM_finality	-1.35	0.26	0.07	< 0.001	***
PRAGM_handing over turn	-1.13	0.32	0.04	< 0.001	***
PRAGM_listing	0.19	1.21	0.04	< 0.001	***
PRAGM_Other (insecurity, checking, emphasis)	-0.30	0.74	0.08	< 0.001	***
INTERVIEWER_INFLUENCE _post-interaction	0.07	1.08	0.03	0.0080	**
TOPIC_entertainment	-0.03	0.97	0.03	0.4235	
TOPIC_experience	0.05	1.05	0.03	0.1510	
TOPIC_self	-0.14	0.87	0.03	< 0.001	***
SEX_male	-0.28	0.75	0.03	< 0.001	***
INT_SEX_male	-0.26	0.77	0.04	< 0.001	***
UPTALK_uptalk	0.69	2.00	0.08	< 0.001	***
QUOTED_quoted	0.50	1.66	0.10	< 0.001	***
REGION_SAmE:POSITION_edge	0.72	2.06	0.08	< 0.001	***
REGION_AmEO:POSITION_edge	0.36	1.44	0.10	< 0.001	***
REGION_CzE:POSITION_edge	1.75	5.76	0.07	< 0.001	***
REGION_GerE:POSITION_edge	1.91	6.75	0.07	< 0.001	***
REGION_MadE:POSITION_edge	0.88	2.41	0.08	< 0.001	***
REGION_MurE:POSITION_edge	0.51	1.67	0.09	< 0.001	***
SEX_male:INT_SEX_male	0.26	1.30	0.08	0.0018	**

Table 55. Descriptive statistics for pitch accent frequency (median and IQR) normalized in per hundred words (absolute tone frequency/number of words by speaker*100) by REGION and SEX, median and IQR frequencies of 0 were deleted from the table

REGION	SEX	TONE_LABEL	Median	IQR
SAmE	female	!H*	0.00	0.18
CzE	female	!H*	0.54	0.70
GerE	female	!H*	0.18	0.35
CzE	male	!H*	0.44	0.69
GerE	male	!H*	0.17	0.18
SAmE	female	H+L*	0.00	0.09
GerE	female	H+L*	0.09	0.35
MadE	female	H+L*	0.18	0.35
MurE	female	H+L*	0.00	0.35
AmEO	male	H+L*	0.00	0.14
CzE	male	H+L*	0.09	0.18
GerE	male	H+L*	0.18	0.37
MadE	male	H+L*	0.18	0.18
MurE	male	H+L*	0.00	0.14
NBrE	female	H*	12.89	2.79
SAmE	female	H*	10.54	3.63
AmEO	female	H*	11.02	0.48
CzE	female	H*	12.30	3.72
GerE	female	H*	10.98	3.32
MadE	female	H*	14.16	1.72
MurE	female	H*	16.56	2.49
NBrE	male	H*	11.49	4.25
SAmE	male	H*	10.31	4.71
AmEO	male	H*	10.30	3.26
CzE	male	H*	11.24	3.76
GerE	male	H*	8.67	3.24
MadE	male	H*	11.37	2.57
MurE	male	H*	17.24	2.26
AmEO	female	L+H*	0.00	0.13
AmEO	male	L+H*	0.00	0.18
NBrE	female	L*	1.62	2.05
SAmE	female	L*	6.87	3.46
AmEO	female	L*	4.10	1.28
CzE	female	L*	4.38	2.21
GerE	female	L*	6.80	2.85
MadE	female	L*	2.83	2.22
MurE	female	L*	3.90	1.96
NBrE	male	L*	4.14	2.67
SAmE	male	L*	3.65	3.27
AmEO	male	L*	3.98	4.02
CzE	male	L*	3.99	2.75

REGION	SEX	TONE_LABEL	Median	IQR
GerE	male	L*	7.09	1.97
MadE	male	L*	5.13	1.03
MurE	male	L*	2.33	0.30
CzE	female	L*+H	0.00	0.18
NBrE	female	^H*	0.52	0.56
SAmE	female	^H*	0.88	0.43
AmEO	female	^H*	0.27	0.57
CzE	female	^H*	0.73	1.01
GerE	female	^H*	0.72	0.61
MadE	female	^H*	0.18	0.26
MurE	female	^H*	0.40	0.79
NBrE	male	^H*	0.00	0.04
SAmE	male	^H*	0.18	0.49
AmEO	male	^H*	0.00	0.18
CzE	male	^H*	1.22	1.13
GerE	male	^H*	0.18	0.35
MadE	male	^H*	0.18	0.18

Table 56. Descriptive statistics for phrase accent frequency (median and IQR) normalized in per hundred words (absolute tone frequency/number of words by speaker*100) by REGION and SEX, median and IQR frequencies of 0 were deleted from the table

REGION	SEX	TONE_LABEL	Median	IQR
NBrE	female	!H-	0.54	0.52
SAmE	female	!H-	0.34	0.72
AmEO	female	!H-	0.35	0.14
CzE	female	!H-	1.45	1.26
GerE	female	!H-	1.33	0.74
MadE	female	!H-	0.90	0.54
MurE	female	!H-	0.40	1.55
NBrE	male	!H-	0.27	0.36
SAmE	male	!H-	0.00	0.30
AmEO	male	!H-	0.00	0.14
CzE	male	!H-	1.42	0.13
GerE	male	!H-	0.71	0.90
MadE	male	!H-	0.53	0.88
MurE	male	!H-	0.54	0.41
NBrE	female	H-	3.41	1.58
SAmE	female	H-	3.16	1.87
AmEO	female	H-	3.23	0.24
CzE	female	H-	4.00	1.41
GerE	female	H-	4.33	2.59
MadE	female	H-	4.40	1.86
MurE	female	H-	5.57	1.72

REGION	SEX	TONE_LABEL	Median	IQR
NBrE	male	H-	2.86	1.35
SAmE	male	H-	2.84	0.89
AmEO	male	H-	1.72	1.32
CzE	male	H-	3.58	2.06
GerE	male	H-	3.77	2.01
MadE	male	H-	3.97	3.89
MurE	male	H-	5.32	2.29
NBrE	female	L-	10.62	1.99
SAmE	female	L-	13.49	3.53
AmEO	female	L-	9.68	0.99
CzE	female	L-	5.61	3.30
GerE	female	L-	8.70	3.36
MadE	female	L-	9.94	2.36
MurE	female	L-	10.51	3.14
NBrE	male	L-	11.51	2.07
SAmE	male	L-	11.37	2.05
AmEO	male	L-	11.49	1.89
CzE	male	L-	5.72	0.82
GerE	male	L-	9.01	2.17
MadE	male	L-	11.17	1.75
MurE	male	L-	9.89	3.64
NBrE	female	^H-	0.00	0.18
SAmE	female	^H-	0.18	0.80
AmEO	female	^H-	0.18	0.27
CzE	female	^H-	1.10	1.13
GerE	female	^H-	0.73	0.75
MadE	female	^H-	0.71	0.87
MurE	female	^H-	1.24	1.67
SAmE	male	^H-	0.00	0.14
CzE	male	^H-	0.80	0.47
GerE	male	^H-	0.18	0.01
MadE	male	^H-	0.35	0.18
MurE	male	^H-	0.18	0.13

Table 57. Descriptive statistics for initial and final boundary tone frequency (median and IQR) normalized in per hundred words (absolute tone frequency/number of words by speaker*100) by REGION and SEX, median and IQR frequencies of 0 were deleted from the table

REGION	SEX	TONE_LABEL	Median	IQR
SAmE	female	!H%	0.00	0.18
AmEO	female	!H%	0.09	0.18
CzE	female	!H%	0.49	0.37
GerE	female	!H%	0.18	0.42

REGION	SEX	TONE_LABEL	Median	IQR
MadE	female	!H%	0.00	0.18
MurE	female	!H%	0.00	0.31
NBrE	male	!H%	0.00	0.18
SAmE	male	!H%	0.00	0.14
CzE	male	!H%	0.44	0.43
GerE	male	!H%	0.18	0.37
MadE	male	!H%	0.35	0.18
MurE	male	!H%	0.00	0.14
NBrE	female	%H	0.18	0.18
SAmE	female	%H	0.00	0.34
AmEO	female	%H	0.18	0.28
CzE	female	%H	0.00	0.34
GerE	female	%H	0.18	0.36
MadE	female	%H	0.00	0.22
MurE	female	%H	0.18	0.36
NBrE	male	%H	0.00	0.22
AmEO	male	%H	0.00	0.14
CzE	male	%H	0.18	0.36
GerE	male	%H	0.00	0.18
MadE	male	%H	0.00	0.18
MurE	male	%H	0.09	0.45
NBrE	female	H%	0.53	0.71
SAmE	female	H%	0.56	0.61
AmEO	female	H%	0.62	0.30
CzE	female	H%	0.73	0.52
GerE	female	H%	1.10	1.05
MadE	female	H%	0.90	0.46
MurE	female	H%	1.18	1.04
NBrE	male	H%	0.53	0.42
SAmE	male	H%	0.37	0.32
AmEO	male	H%	0.36	0.85
CzE	male	H%	0.54	0.25
GerE	male	H%	1.08	1.09
MadE	male	H%	1.06	0.52
MurE	male	H%	1.34	0.46
NBrE	female	L%	4.40	1.67
SAmE	female	L%	3.78	1.16
AmEO	female	L%	3.66	1.08
CzE	female	L%	2.36	1.38
GerE	female	L%	3.51	2.22
MadE	female	L%	3.98	1.95
MurE	female	L%	4.76	2.80
NBrE	male	L%	5.32	1.48
SAmE	male	L%	4.78	1.32
AmEO	male	L%	3.98	2.04

REGION	SEX	TONE_LABEL	Median	IQR
CzE	male	L%	1.90	1.21
GerE	male	L%	3.78	1.03
MadE	male	L%	3.89	0.96
MurE	male	L%	3.90	0.80
AmEO	female	^H%	0.00	0.13
CzE	female	^H%	0.68	0.73
GerE	female	^H%	1.06	0.93
MadE	female	^H%	0.18	0.35
MurE	female	^H%	0.35	0.48
CzE	male	^H%	0.43	0.98
GerE	male	^H%	0.18	0.53
MurE	male	^H%	0.18	0.00

Table 58. Tunes by ip in relative frequencies (%)[87] by REGION and SEX (excluding the complex tune patterns)

REGION	SEX	HL	HH	LF	LL	HF	LH	HLH	LHL	Total ips
CzE	**F**	30.75	**31.15**	4.65	7.47	**8.97**	**9.77**	5.12	2.12	2,732
GerE		29.45	**22.35**	9.94	13.44	**9.22**	**10.83**	3.15	1.61	2,917
MadE		45.05	**24.70**	10.06	6.30	**8.88**	3.07	1.51	0.43	1,858
MurE		49.96	**21.33**	5.71	5.08	**8.88**	5.31	1.59	2.14	1,261
NBrE		**66.19**	14.01	3.81	6.85	5.43	1.62	1.17	0.91	1,970
SAmE		44.41	10.52	**17.19**	13.18	3.36	**6.51**	1.52	3.31	1,844
AmEO		50.71	15.42	9.13	8.11	4.46	**7.10**	2.43	2.64	493
CzE	**M**	32.94	24.07	3.50	9.11	16.12	**7.71**	3.97	2.57	428
GerE		26.01	10.74	**25.89**	12.33	13.80	**7.81**	1.71	1.71	819
MadE		35.53	15.53	**21.70**	6.38	14.68	4.47	1.06	0.64	470
MurE		51.36	18.69	7.08	2.36	14.70	2.00	2.18	1.63	551
NBrE		52.69	8.53	**16.15**	11.86	5.81	2.54	0.73	1.69	1,653
SAmE		58.15	4.39	**17.54**	5.76	6.14	4.01	1.13	2.88	798
AmEO		58.80	2.05	**21.28**	4.64	4.09	4.37	2.86	1.91	733
Total		8,109	3,335	2,044	1,654	1,489	1,129	433	334	18,527

[87]Relative frequencies are presented here for ease of comprehension and comparison across sexes; statistical tests were conducted on the basis of absolute frequencies.

Table 59. Tunes in final position by REGION and SEX in relative frequencies

REGION	SEX	HL	HH	LF	LL	HF	LH	HLH	LHL	Total ips
CzE	**F**	32.98	27.11	3.31	13.87	3.31	11.95	4.91	2.56	937
GerE		28.94	17.41	7.96	17.99	4.89	16.33	4.48	1.99	1206
MadE		48.34	15.73	13.91	11.42	3.48	4.30	2.15	0.66	604
MurE		53.73	14.91	6.58	9.43	5.48	4.39	2.63	2.85	456
NBrE		70.58	5.36	3.91	12.03	2.17	1.88	2.75	1.30	690
SAmE		35.59	6.38	18.38	20.50	0.58	7.93	3.87	6.77	517
AmEO		50.33	6.54	9.15	12.42	2.61	7.84	4.58	6.54	153
CzE	**M**	37.24	17.24	3.45	15.17	12.41	10.34	2.76	1.38	145
GerE		24.38	10.94	24.38	17.19	7.50	10.94	2.19	2.50	320
MadE		31.41	13.46	27.56	10.26	7.69	5.13	3.21	1.28	156
MurE		54.44	13.89	10.00	4.44	8.33	1.67	6.67	0.56	180
NBrE		50.92	5.37	16.87	18.10	1.99	2.61	1.53	2.61	652
SAmE		56.52	1.67	19.06	9.03	2.68	3.34	3.01	4.68	299
AmEO		52.19	2.19	21.90	8.03	0.36	6.20	6.57	2.55	274
Total ips		2,866	859	748	935	249	526	236	170	6,589

Table 60. Tunes in non-final position by REGION and SEX in relative frequencies

REGION	SEX	HL	HH	LF	LL	HF	LH	HLH	LHL	Total ips
CzE	**F**	29.58	33.26	5.35	4.12	11.92	8.64	5.24	1.89	1,795
GerE		29.81	25.83	11.34	10.23	12.27	6.95	2.22	1.34	1,711
MadE		43.46	29.03	8.21	3.83	11.48	2.47	1.20	0.32	1,254
MurE		47.83	24.97	5.22	2.61	10.81	5.84	0.99	1.74	805
NBrE		63.83	18.67	3.75	4.06	7.19	1.48	0.31	0.70	1,280
SAmE		47.85	12.13	16.73	10.32	4.45	5.95	0.60	1.96	1,327
AmEO		50.88	19.41	9.12	6.18	5.29	6.76	1.47	0.88	340
CzE	**M**	30.74	27.56	3.53	6.01	18.02	6.36	4.59	3.18	283
GerE		27.05	10.62	26.85	9.22	17.84	5.81	1.40	1.20	499
MadE		37.58	16.56	18.79	4.46	18.15	4.14	0.00	0.32	314
MurE		49.87	21.02	5.66	1.35	17.79	2.16	0.00	2.16	371
NBrE		53.85	10.59	15.68	7.79	8.29	2.50	0.20	1.10	1,001
SAmE		59.12	6.01	16.63	3.81	8.22	4.41	0.00	1.80	499
AmEO		62.75	1.96	20.92	2.61	6.32	3.27	0.65	1.53	459
Total		5,243	2,476	1,296	719	1,240	603	197	164	11,938

Table 61. Tunes in relative frequencies (%) of ips by REGION, SEX, PATTERN, and POSITION (final vs. non-final)

		IP-final		IP-non-final		IP-final		Total	
REGION	**SEX**	**H* H- L%**	**H* L- L%**	**H* L-**	**L* H-**	**L* H- H%**	**L* L- H%**	**Final**	**Non_ final**
NBrE	**F**	16.23	54.35	63.83	1.48	0.29	1.59	690	1280
SAmE		6.19	29.40	47.85	5.95	3.87	4.06	517	1327
AmEO		7.84	42.48	50.88	6.76	4.58	3.27	153	340
CZE		7.68	25.29	29.58	8.64	9.50	2.45	937	1795
GerE		5.31	23.63	29.81	6.95	11.53	4.81	1206	1711
MadE		7.62	40.73	43.46	2.47	3.31	0.99	604	1254
MurE		16.89	36.84	47.83	5.84	2.85	1.54	456	805
NBrE	**M**	11.04	39.88	53.85	2.50	1.23	1.38	652	1001
SAmE		17.39	39.13	59.12	4.41	2.01	1.34	299	499
AmEO		11.31	40.88	62.75	3.27	2.92	3.28	274	459
CZE		4.83	32.41	30.74	6.36	6.90	3.45	145	283
GerE		6.88	17.50	27.05	5.81	8.44	2.50	320	499
MadE		7.69	23.72	37.58	4.14	3.85	1.28	156	314
MurE		14.44	40.00	49.87	2.16	1.11	0.56	180	371

Table 62. Results for mixed effects model with f0 level in (f0 mean) Hz. Final model: F0_LEVEL_MEAN ~ TUNE_PATTERN_confl * REGION + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + TOPIC_INITIATIONS + LENGTH + HP_NORM + (1 | SPKR)

Fixed effects:	**Estimates**	**Std. Error**	**t-value**	**p-value**	
(Intercept)	202.7068	4.3568	46.53	< 0.001	***
PATTERN_HIGH_HIGH	10.8096	0.8752	12.35	< 0.001	***
PATTERN_LOW_HIGH	-4.5722	1.8313	-2.5	0.0125	*
PATTERN_LOW_LOW	-15.9875	0.9037	-17.69	< 0.001	***
REGION_SAmE	11.3598	4.6519	2.44	0.0154	*
REGION_AmEO	6.0459	5.5718	1.09	0.279	
REGION_CzE	4.364	4.4818	0.97	0.3312	
REGION_GerE	7.3101	4.1944	1.74	0.0827	.
REGION_MadE	9.4342	5.2645	1.79	0.0744	.
REGION_MurE	7.8086	6.4807	1.2	0.2295	
SEX_male	-80.7474	2.8793	-28.04	< 0.001	***
SPEAKING_STYLE_dia	-0.4414	0.4735	-0.93	0.3512	
POSITION_FINAL	-6.1215	0.2975	-20.58	< 0.001	***
PRAGM_checking	12.3931	1.6236	7.63	< 0.001	***
PRAGM_emphasis	7.9077	2.8933	2.73	0.0063	**
PRAGM_finality	-1.9381	0.9301	-2.08	0.0372	*
PRAGM_handing over turn	-3.9595	0.5989	-6.61	< 0.001	***
PRAGM_insecurity	-0.7029	1.7264	-0.41	0.6839	
PRAGM_listing	-5.4216	0.5355	-10.12	< 0.001	***
INTERVIEWER_INFLUENCE_post-interaction	6.0978	0.3954	15.42	< 0.001	***
TOPIC_INITIATIONS	-1.7676	0.859	-2.06	0.0408	*
LENGTH	0.7038	0.1187	5.93	< 0.001	***
HP_NORM	-0.9474	0.6031	-1.57	0.1176	
PATTERN_HIGH_HIGH:REGION_SAmE	1.2411	1.4574	0.85	0.3944	

Fixed effects:	Estimates	Std. Error	t-value	p-value	
PATTERN_LOW_HIGH:REGION_SAmE	2.9246	2.3394	1.25	0.2113	
PATTERN_LOW_LOW:REGION_SAmE	-6.9075	1.2516	-5.52	< 0.001	***
PATTERN_HIGH_HIGH:REGION_AmEO	0.9246	2.0102	0.46	0.6456	
PATTERN_LOW_HIGH:REGION_AmEO	4.9971	2.6994	1.85	0.0642	.
PATTERN_LOW_LOW:REGION_AmEO	-0.5048	1.6262	-0.31	0.7563	
PATTERN_HIGH_HIGH:REGION_CzE	7.2563	1.1805	6.15	< 0.001	***
PATTERN_LOW_HIGH:REGIONd_ CzE	12.9705	2.1145	6.13	< 0.001	***
PATTERN_LOW_LOW:REGIONd_ CzE	0.2937	1.4405	0.2	0.8385	
PATTERN_HIGH_HIGH:REGION_GerE	5.0703	1.1886	4.27	< 0.001	***
PATTERN_LOW_HIGH:REGION_ GerE	12.4283	2.0997	5.92	< 0.001	***
PATTERN_LOW_LOW:REGION_ GerE	0.0318	1.2371	0.03	0.9795	
PATTERN_HIGH_HIGH:REGION_MadE	8.3503	1.244	6.71	< 0.001	***
PATTERN_LOW_HIGH:REGION_ MadE	12.6895	2.5822	4.91	< 0.001	***
PATTERN_LOW_LOW:REGION_ MadE	1.8731	1.3899	1.35	0.1778	
PATTERN_HIGH_HIGH:REGION_ MurE	4.2697	1.3299	3.21	0.0013	**
PATTERN_LOW_HIGH:REGION_ MurE	17.0915	2.5995	6.57	< 0.001	***
PATTERN_LOW_LOW:REGION_ MurE	4.9138	1.733	2.84	0.0046	**

Number of obs.=17,566, Random intercepts: Speaker SD (18.84), Residual SD (17.9); Log likelihood =-76104, Conditional $R^{2=}$0.86.

Table 63. Results for mixed effects model with f0 span in STs. Final model: F0_SPAN_ST ~ TUNE_PATTERN_confl * REGION + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + LENGTH + HP_NORM + TOPIC_INITIATIONS + (1 | SPKR)

Fixed effects:	Estimates	Std. Error	t-value	p-value	
(Intercept)	5.0539	0.2528	19.99	< 0.001	***
PATTERN_confl_HIGH_HIGH	-2.7198	0.1556	-17.48	< 0.001	***
PATTERN_confl_LOW_HIGH	-1.4919	0.3284	-4.54	< 0.001	***
PATTERN_confl_LOW_LOW	-2.269	0.1633	-13.89	< 0.001	***
REGION _SAME	-0.1939	0.2703	-0.72	0.47377	
REGION _AMEO	0.1734	0.3255	0.53	0.59453	
REGION_CzE	-0.678	0.2691	-2.52	0.01226	*
REGION_GerE	-0.8217	0.2524	-3.26	0.00126	**
REGION_MadE	-1.8108	0.3068	-5.9	< 0.001	***
REGION_MurE	-1.9116	0.3712	-5.15	< 0.001	***
SEX_male	-1.0819	0.1617	-6.69	< 0.001	***
SPEAKING_STYLE_dia	0.1795	0.08	2.24	0.02483	*
POSITION_FINAL	0.7847	0.0537	14.6	< 0.001	***
PRAGM_checking	1.1236	0.2888	3.89	< 0.001	***
PRAGM_emphasis	1.1156	0.5032	2.22	0.02664	*
PRAGM_finality	0.0989	0.1698	0.58	0.56031	
PRAGM_handing over turn	0.4268	0.1083	3.94	< 0.001	***
PRAGM_insecurity	0.3059	0.3095	0.99	0.32286	
PRAGM_listing	-0.5224	0.0967	-5.4	< 0.001	***
INTERVIEWER_INFLUENCE_post-interaction	0.454	0.0704	6.45	< 0.001	***
LENGTH	0.2873	0.0213	13.5	< 0.001	***
HP_NORM	-0.0168	0.0338	-0.5	0.6192	
TOPIC_INITIATIONS	0.0026	0.049	0.05	0.95767	
PATTERN_conflHIGH_HIGH:REGION_SAME	-0.0424	0.2562	-0.17	0.86846	
PATTERN_conflLOW_HIGH:REGION_SAME	0.0541	0.4185	0.13	0.89718	
PATTERN_conflLOW_LOW:REGION_SAME	-1.0609	0.2268	-4.68	< 0.001	***
PATTERN_conflHIGH_HIGH:REGION_AMEO	-0.0398	0.3551	-0.11	0.91076	
PATTERN_conflLOW_HIGH:REGION_AMEO	-0.44	0.4889	-0.9	0.36816	
PATTERN_conflLOW_LOW:REGION_AMEO	-1.2897	0.297	-4.34	< 0.001	***
PATTERN_conflHIGH_HIGH:REGION_CzE	0.8617	0.2097	4.11	< 0.001	***

Fixed effects:	Estimates	Std. Error	t-value	p-value	
PATTERN_conflLOW_HIGH:REGION_ CzE	0.9618	0.3793	2.54	0.01124	*
PATTERN_conflLOW_LOW:REGION_ CzE	0.2483	0.2597	0.96	0.33893	
PATTERN_conflHIGH_HIGH:REGION_GerE	0.5675	0.2106	2.69	0.00705	**
PATTERN_conflLOW_HIGH:REGION_GerE	1.2976	0.3754	3.46	< 0.001	***
PATTERN_conflLOW_LOW:REGION_ GerE	0.3086	0.222	1.39	0.16463	
PATTERN_conflHIGH_HIGH:REGION_ MadE	1.5641	0.2236	7	< 0.001	***
PATTERN_conflLOW_HIGH:REGION_MadE	1.9643	0.4658	4.22	< 0.001	***
PATTERN_conflLOW_LOW:REGION_MadE	0.1993	0.2617	0.76	0.44624	
PATTERN_conflHIGH_HIGH:REGION_MurE	1.9333	0.2379	8.13	< 0.001	***
PATTERN_conflLOW_HIGH:REGION_MurE	1.8937	0.4677	4.05	< 0.001	***
PATTERN_conflLOW_LOW:REGION_MurE	0.7479	0.3194	2.34	0.01922	*

Number of obs.=16,220, Random intercepts: Speaker SD (0.87), Residual SD (1.95); Log likelihood =-34172, Conditional $R^{2=}$0.38.

Table 64. Results for mixed effects model with f0 span in STs for the complete learner data only. Final model: log2(F0_SPAN_ST) ~ TUNE_PATTERN_confl * REGION * STAB + SEX + SPEAKING_STYLE + POSITION + PRAGM + INTERVIEWER_INFLUENCE + LENGTH + TOPIC_INITIATIONS + YOEAS + YOEAU + NO_OFL + (1 | SPKR)

Fixed effects	Estimates	Std. Error	t-value	p-value	
(Intercept)	1.517012	0.308301	4.92	< 0.001	***
TUNE_PATTERN_confl_HH	-0.896802	0.184604	-4.86	< 0.001	***
TUNE_PATTERN_confl_LH	-0.022358	0.224448	-0.1	0.921	
TUNE_PATTERN_confl_LL	-1.218511	0.266107	-4.58	< 0.001	***
REGION_GerE	-0.203714	0.24948	-0.82	0.415	
REGION_MadE	-0.413589	0.262577	-1.58	0.117	
STAB_NO_STAB	-0.391439	0.243839	-1.61	0.11	
STAB_SHORT	-0.259308	0.205838	-1.26	0.209	
SEX_xmale	-0.369889	0.087779	-4.21	< 0.001	***
SPEAKING_STYLE_dia	0.03675	0.041584	0.88	0.377	
POSITION_final	0.515436	0.032086	16.06	< 0.001	***
PRAGM_CHECK	0.49056	0.124178	3.95	< 0.001	***
PRAGM_EMP	0.425497	0.234989	1.81	0.07	.
PRAGM_FIN	0.003903	0.084493	0.05	0.963	
PRAGM_HOT	-0.011391	0.070468	-0.16	0.872	
PRAGM_INSEC	0.092567	0.137373	0.67	0.5	
PRAGM_LIST	-0.333425	0.055987	-5.96	< 0.001	***
INTERVIEWER_INFLUENCE_post-interaction	0.201986	0.044109	4.58	< 0.001	***
LENGTH	0.129086	0.011023	11.71	< 0.001	***
TOPIC_INITIATIONS	-0.000764	0.023547	-0.03	0.974	
YOEAS	0.000729	0.016931	0.04	0.966	
YOEAU	0.047321	0.030077	1.57	0.118	
NO_OFL	-0.007635	0.039807	-0.19	0.848	
TUNE_PATTERN_confl_HH:REGION_GerE	-0.084152	0.242212	-0.35	0.728	
TUNE_PATTERN_confl_LH:REGION_GerE	-0.585064	0.302055	-1.94	0.053	.
TUNE_PATTERN_confl_LL:REGION_GerE	-0.076667	0.3057	-0.25	0.802	
TUNE_PATTERN_confl_HH:REGION_MadE	0.112867	0.238562	0.47	0.636	
TUNE_PATTERN_confl_LH:REGION_MadE	0.332188	0.38432	0.86	0.387	
TUNE_PATTERN_confl_LL:REGION_MadE	-0.409526	0.328609	-1.25	0.213	
TUNE_PATTERN_confl_HH:STAB_NO_STAB	0.280273	0.226998	1.23	0.217	
TUNE_PATTERN_confl_LH:STAB_NO_STAB	-0.158674	0.297765	-0.53	0.594	
TUNE_PATTERN_confl_LL:STAB_NO_STAB	0.212998	0.359054	0.59	0.553	
TUNE_PATTERN_confl_HH:STAB_SHORT	0.067358	0.197746	0.34	0.733	
TUNE_PATTERN_confl_LH:STAB_SHORT	-0.153166	0.242514	-0.63	0.528	
TUNE_PATTERN_confl_LL:STAB_SHORT	0.155262	0.283184	0.55	0.584	
REGION_GerE:STAB_NO_STAB	0.406909	0.31765	1.28	0.201	

Fixed effects	Estimates	Std. Error	t-value	p-value	
REGION_MadE:STAB_NO_STAB	0.357669	0.347928	1.03	0.305	
REGION_GerE:STAB_SHORT	0.036363	0.273972	0.13	0.895	
REGION_MadE:STAB_SHORT	0.201438	0.289044	0.7	0.487	
TUNE_HH:REGION_GerE:STABNO_STAB	-0.301894	0.299517	-1.01	0.314	
TUNE_LH:REGION_GerE:STABNO_STAB	0.693403	0.389613	1.78	0.075	.
TUNE_LL:REGION_GerE:STABNO_STAB	-0.145985	0.407314	-0.36	0.72	
TUNE_HH:REGION_MadE:STABNO_STAB	-0.44079	0.314708	-1.4	0.161	
TUNE_LH:REGION_MadE:STABNO_STAB	-0.594052	0.528576	-1.12	0.261	
TUNE_LL:REGION_MadE:STABNO_STAB	-0.222943	0.444684	-0.5	0.616	
TUNE_HH:REGION_GerE:STABSHORT	-0.085761	0.264363	-0.32	0.746	
TUNE_LH:REGION_GerE:STABSHORT	0.776038	0.331246	2.34	0.019	*
TUNE_LL:REGION_GerE:STABSHORT	-0.045345	0.332116	-0.14	0.891	
TUNE_HH:REGION_MadE:STABSHORT	0.017631	0.263569	0.07	0.947	
TUNE_LH:REGION_MadE:STABSHORT	0.073085	0.434716	0.17	0.866	
TUNE_LL:REGION_MadE:STABSHORT	-0.056877	0.361456	-0.16	0.875	

Number of obs.=9,034, Random intercepts: Speaker SD (0.316), Residual SD (1.383); Log likelihood =-15840, Conditional $R^{2=}0.23$.

Table 65. Frequent uptalkers with more than three uptalk tunes of the 560 word samples and their tune choice and pragmatic functions used

SPKR	LH	LHH	HH	HLH	CHECK	CON	INSEC	LIST	Total
CZ007	6	3	0	0	1	8	0	0	9
CZ028	4	0	0	0	1	3	0	0	4
CZ030	3	0	3	0	0	5	0	1	6
CZ032	4	1	1	2	5	1	0	2	8
CZ038	5	1	0	2	5	2	1	0	8
CZ042	4	3	2	1	8	2	0	0	10
CZ046	3	2	1	0	5	1	0	0	6
CZ048	10	5	0	0	14	1	0	0	15
CZ050	1	0	0	3	2	2	0	0	4
GE001	2	2	1	0	5	0	0	0	5
GE003	2	1	1	0	2	2	0	0	4
GE005	2	2	1	0	2	3	0	0	5
GE009	5	2	1	0	5	3	0	0	8
GE012	2	1	4	0	2	5	0	0	7
GE015	8	2	3	0	5	6	0	2	13
GE026	6	5	3	2	4	12	0	0	16
GE027	3	1	0	0	0	3	0	1	4
GE028	5	8	1	0	2	12	0	0	14
GE029	3	2	4	0	1	6	0	2	9
GE032	5	4	0	0	0	9	0	0	9
GE035	1	2	2	0	1	4	0	0	5
GE037	1	2	4	0	3	4	0	0	7
GE038	1	2	3	0	0	6	0	0	6
GE041	3	3	0	0	1	4	0	1	6
GE042	3	1	3	0	0	7	0	0	7
GE043	3	1	0	0	3	1	0	0	4

SPKR	LH	LHH	HH	HLH	CHECK	CON	INSEC	LIST	Total
GE045	3	2	7	0	3	9	0	0	12
GE047	3	1	0	0	0	4	0	0	4
SP032	1	0	5	0	0	6	0	0	6
SP046	5	0	2	0	0	6	0	1	7
Total	107	59	52	10	80	137	1	10	

Table 66. Relative Slope of Rise in STs, absolute Slope of Rise in STs, and Rise Duration in seconds

		RELATIVE_SLOPE_ST				SLOPE_of_RISE_ST				RISE_DURATION_S			
REGION	**SEX**	**MEAN**	**SD**	**MEDIAN**	**IQR**	**MEAN**	**SD**	**MEDIAN**	**IQR**	**MEAN**	**SD**	**MEDIAN**	**IQR**
CzE	female	17.32	12.62	14.47	16.12	15.41	9.41	13.06	11.56	0.46	0.20	0.40	0.21
GerE	female	13.18	12.25	9.83	8.60	17.21	10.03	14.29	12.22	0.46	0.16	0.43	0.18
MadE	female	21.40	13.25	17.59	12.17	19.58	8.78	17.33	12.12	0.35	0.08	0.35	0.09
MurE	female	15.95	9.41	14.77	5.25	23.44	8.71	24.43	12.15	0.32	0.10	0.31	0.16
New-Engl.	female	13.31	14.51	7.62	10.26	11.06	5.26	11.62	7.27	0.44	0.06	0.43	0.08
South	female	5.10	5.48	3.54	3.96	4.19	3.36	3.30	2.45	0.58	0.29	0.50	0.15
West	female	5.27	4.60	5.27	3.26	7.85	4.84	7.85	3.43	0.92	0.10	0.92	0.07
CzE	male	12.96	9.38	8.22	11.62	15.51	8.39	12.57	10.30	0.42	0.20	0.42	0.34
GerE	male	14.43	13.15	11.51	11.35	17.89	17.32	13.31	7.93	0.45	0.15	0.42	0.08
South	male	9.05	1.12	9.05	0.80	18.43	2.09	18.43	1.48	0.40	0.06	0.40	0.04

Table 67. Relative Slope of Rise in STs, absolute Slope of Rise in STs, and Rise Duration in seconds for frequent uptalkers (more than three instances per speaker, female speakers only)

	RELATIVE_SLOPE_ST				SLOPE_of_RISE_ST				RISE_DURATION_S			
REGION	**MEAN**	**SD**	**MEDIAN**	**IQR**	**MEAN**	**SD**	**MEDIAN**	**IQR**	**MEAN**	**SD**	**MEDIAN**	**IQR**
CzE	16.19	10.64	14.19	15.5	14.52	8.65	12.17	7.67	0.45	0.16	0.40	0.19
GerE	12.68	11.73	9.79	8.17	17.27	10.29	14.19	12.25	0.47	0.16	0.43	0.18

Appendix 4. German Summary

Kapitel 1

Einleitung und Überblick

Das zentrale Ziel des vorliegenden Dissertationsprojekts war es, die Sprachproduktion fortgeschrittener Lerner des Englischen in zweierlei Sprachformen – natürlichsprachlichen Monologen sowie Interaktionen – mit der englischer Muttersprachler zu vergleichen, um näher zu beleuchten, wie sich die Prosodie der Lerner von muttersprachlich produzierter Prosodie unterscheidet und inwiefern die beobachteten prosodischen Merkmale von verschiedenen Erstsprachengruppen geteilt werden. Dabei lag der Fokus der Studie auf prosodischen Charakteristika deklarativer Äußerungen, welche aus der spontanen Produktion der Zweitsprache (L2) bzw., im Falle der Kontrollgruppen, der Muttersprache hervorgehen, sowie auf dem Einfluss von Interviewerfragen auf einige der deklarativen Äußerungen. Insgesamt wurden die prosodischen Muster von 135 fortgeschrittenen Englischlernern mit deutschem, tschechischem, und spanischem Hintergrund analysiert, mit dem Ziel, die akustischen Eigenschaften der Grundfrequenz (*f0*) bei Verwendung der Zweitsprache zu identifizieren. Die Performanz der drei Lernergruppen wurde mit der von einsprachigen Muttersprachlern des Britischen (n = 45) und Amerikanischen (n = 45) verglichen.[88] Im Rahmen des autosegmental-metrischen (AM) Ansatzes (Beckman und Pierrehumbert 1986) wurde die Produktion von Prosodie in verschiedenen Tonereignissen untersucht: Tonhöhenakzente, Phrasen- und Grenztöne und deren Grundfrequenz. Um eine phonetische und phonologische Beschreibung dieser tonalen Ereignisse zu erhalten, war eine Untersuchung des Bereichs der Intonationsphrasierung (Intonationsphrasen (IPs) und Intermediärphrasen (ips)) und der Strukturierung dieser Töne zu Sprachmelodien erforderlich. Um schließlich auch die Beziehung zwischen prosodischer Form und sozio-pragmatischer Bedeutung bestimmen zu können, wurde auch eine Analyse der Form-Funktions-Korrelationen durchgeführt.

Gemäß des *Gemeinsamen Europäischen Referenzrahmen für Sprachen* (GER) wird von einem Lerner auf C2-Niveau erwartet, prosodische Merkmale wie Betonung, Rhythmus und Intonation so nutzbar zu machen, dass "finer shades of meaning" (Council of Europe 2018: 136), also feinere Bedeutungsnuancen, zum Ausdruck gebracht werden können. Damit

---

[88] Bitte beachten Sie, dass die Performanz der Sprecher im Falle des spanischen sowie des amerikanischen Korpus große regionale Differenzen aufwies, weshalb eine regionale Separation in jeweils zwei Untergruppen vorgenommen und schlussendlich von insgesamt sieben Varietäten ausgegangen wurde.

sind hauptsächlich pragmatische Funktionen gemeint, welche mit Hilfe prosodischer Merkmale erfüllt werden.

Bei der Prosodie handelt es sich um ein komplexes, schwer greifbares Konzept. Dies ist einerseits den zahlreichen Definitionen, Annotations- und Analysemethoden geschuldet, andererseits jedoch auch dem Umstand, dass sie von verschiedenen, miteinander verzahnten Faktoren beeinflusst wird. Dieses Projekt zielt darauf ab, auf Basis einer kontrastiven korpusbasierten Analyse von Muttersprachler- und Lernerdaten, prosodische Merkmale zu identifizieren, in denen sich Lerner bereits stark an die Muttersprachlernorm angenähert haben bzw. solche, in welchen sie noch verstärkt von ihr abweichen. Die Daten in der vorliegenden Studie wurden pragmatisch, phonologisch und phonetisch analysiert, indem das Auftreten von Intermediärphrasen mit bestimmten pragmatischen Funktionen und deren Tonwahl und Grundfrequenz quantifiziert wurden. Auf Grundlage der Ergebnisse konnten so fremdsprachendidaktisch relevante Implikationen für das Lehren, Lernen und Bewerten der englischen Prosodie gewonnen werden. Somit hatte die Arbeit auch das anwendungsorientierte Ziel, dazu beizutragen, Kompetenzskalen für den GER zu erstellen und einen Beitrag auf dem Gebiet des Fremdsprachenerwerbs und der Sprachbewertung zu leisten.

Im ersten Kapitel wurden zunächst grundsätzliche Konzepte definiert, um sich schließlich im theoretischen Fachdiskurs positionieren zu können, bevor, im Folgenden, die Unterrichtspraktiken – in Bezug auf das Englische im Allgemeinen sowie die Prosodie im Besonderen – in Deutschland, Tschechien und Spanien dargestellt wurden. Dabei machte eine kurze Einordnung des allgemeinen politischen Status der Sprache sowie des Englischlernens innerhalb dieser Länder deutlich, dass die Spracherwerbssituationen durchaus vergleichbar sind. In allen drei Ländern hat Englisch keinen offiziellen Status, jedoch ist die Sprache relevant für Studium, Beruf und Reisen. Was die Prosodie anbelangt, so nimmt diese in den Ländern eine ähnlich vernachlässigte Stellung im Unterricht ein.

Danach wird auf die Relevanz des Konzepts des Muttersprachlers eingegangen, der in der vorliegenden Untersuchung als relevanter Maßstab betrachtet wird, bevor im weiteren Verlauf, eine Zusammenfassung unterschiedlicher Zweitsprachenerwerbstheorien und Sprachmodelle folgt, die hauptsächlich auf der segmentalen Ebene zu verorten sind, wobei der Fokus hier auf Mennen's (2015) *L2 Intonational Learning Theory* (LILt) liegt, welches das erste Zweitsprachenerwerb-basierte Intonationsmodell darstellt. Mennen's Vorhersagen und Forschungsfragen werden auf Grundlage der vier Intonationsdimensionen (systemische, realisierende, verteilende und funktionelle Dimension) des LILt Modells erforscht. Des

Weiteren wird auch die CAT (*Communication Accommodation Theory*; Giles et al. 1991) Theorie überprüft und inwiefern sich die Lerner der Grundfrequenz der entsprechenden Interviewer annähern (*prosodic entrainment)*.

Das Kapitel schließt mit einer Übersicht von früheren Studien zur Prosodie und unterschiedlichen extralinguistischen Faktoren, die den Zweitsprachenerwerbsprozess beeinflussen. Alle diskutierten sozialpsychologischen und kontextuellen Variablen (Aufgabenbedingungen und Interviewerrolle) tragen individuell zum erfolgreichen Erwerb der Zweitsprache bei. Die meisten extralinguistischen Faktoren wie Alter, Geschlecht, Länge des Auslandsaufenthaltes, Motivation sowie Dauer und Art des Unterrichts weisen distinktive Ergebnisse auf und beeinflussen die L2-Produktion in unterschiedlichem Maße. Während einige Studien positive Korrelationen attestieren, zeigen andere keine oder gemischte Effekte. Oftmals wurde das Englische erst nach der sog. „kritischen Phase" (Lenneberg 1967) erlernt, weswegen Lernersprache fast immer durch einen ausgeprägten Akzent und nicht-muttersprachliche Intonationsmuster gekennzeichnet ist. Problematisch hierbei ist, dass viele der Faktoren nur schwer zu messen sind und es dementsprechend ebenso schwierig ist, nachzuweisen, ob ein – und wenn ja welcher – Faktor in welchem Umfang für das Erreichen des Ziels (*ultimate attainment*) "verantwortlich" ist. Selbstverständlich hängt dies stets von der Definition des Ziels ab und davon, auf welche linguistische Ebene man sich bezieht. All diese Faktoren sind miteinander verflochten und können sich gegenseitig beeinflussen. So ist es beispielsweise sehr wahrscheinlich, dass das Geschlecht eines Lerners mit anderen Variablen – wie etwa der höheren Motivation, einer positiveren Einstellung gegenüber den L2-Sprechern oder unterschiedlichen Lernstilen, welche alle insbesondere bei weiblichen Lernerinnen beobachtet wurden – interagiert, die den erfolgreichen Erwerb der L2 bedingen können. Daher wird es als wichtig erachtet, diese extralingustischen Faktoren nicht isoliert zu untersuchen. Stattdessen sollte versucht werden, so viele miteinander verzahnte Variablen wie möglich in die Analyse einzubeziehen, um dem Umstand gerecht zu werden, dass es sich bei dem Erwerb eines ausländischen Akzents um eine äußerst komplexe Thematik handelt. Die soziolinguistischen Variablen in der vorliegenden Studie beschränken sich auf Dialekt, Geschlecht, Alter, Sprachkenntnisse (verschiedene Maßnahmen), Lernkontext (z.B. Auslandsaufenthalte, Anzahl der Jahre des Englischspracherwerbs an der Schule und Universität), sowie andere extralinguistische Variablen zur Interviewsituation (z.B. Sprechsituation/Sprechstil, Geschlecht des Interviewers).

Die Arbeit leistet somit einen wichtigen Beitrag zu der wachsenden korpusbasierten Forschung und verfolgt einen quantitativen Ansatz mit gemischten und multifaktoriellen

Methoden. Dabei werden die phonologische und phonetische Form der Grundfrequenz analysiert, um Variation sowie Universalien im muttersprachlichen und interimssprachlichen, spontanen Sprachgebrauch aufzudecken, die zusätzlich durch extralinguistische und kontextuelle Faktoren bedingt sind. Eine solche Analyse trägt zum Verständnis der prosodischen Variation in der muttersprachlichen und nicht-muttersprachlichen Sprache bei und deckt überdies auf, welche Faktoren eine solche Variation vorhersagen können. Die Studie hat somit Auswirkungen auf die theoretische und angewandte Linguistik, da die Ergebnisse für Bereiche wie theoretische Zweitspracherwerbsprozesse, Sprachbewertung, Sprachunterricht, Entwicklung computergestützter Sprachunterrichtsmaterialien und Sprachtechnologie relevant sind.

Kapitel 2

Prosodie in Englisch als Muttersprache und als Fremdsprache, und Deutsch, Tschechisch, Spanisch als Muttersprache

Im zweiten Kapitel der vorliegenden Dissertation werden zunächst zwei konkurrierende Modelle zur Modellierung und Beschreibung der englischen Intonation vor- und gegenübergestellt, wobei das Tonsequenzmodell von Pierrehumbert (1980) für die Bedürfnisse dieser Arbeit bevorzugt wurde, hauptsächlich, um die Ergebnisse mit früheren Studien vergleichbar zu machen. Das *Tone and Break Indices* (ToBI) System (Silverman et al. 1992), welches auf dem Tonsequenzmodell von Pierrehumbert (1980) basiert, wird im Rahmen der empirischen Studie der vorliegenden Arbeit für die prosodische Analyse der Audiodateien zugrundegelegt. Dabei handelt es sich um ein Intonationsmodell, das als phonologisches Transkriptionssystem für die Intonation des Englischen dient. Da sich die Arbeit der Analyse von Fremdsprachenlernern aus drei typologisch unterschiedlichen Erstsprachen widmet, wurde im darauffolgenden Kapitel ein allgemeiner Vergleich der deutschen, tschechischen, spanischen und englischen Prosodie angestellt, um voraussagen zu können, in welchen Bereichen möglicherweise Probleme für die entsprechenden L2-Sprechergruppen zu erwarten sind. Eine stark abweichende Performanz aller Lernergruppen von der Muttersprachlernorm auf Grund eines negativen Transfers ist beispielsweise im Grundfrequenzumfang zu erwarten, während für das Toninventar mit positivem Transfer zu rechnen ist, da dieses in allen Erstsprachen deutliche Ähnlichkeiten aufweist. Zusätzlich kann, im Falle des Tschechischen sowie des Spanischen, auf Grund des eher silbengesteuerten Rhythmus, von einem negativen Transfer bezüglich der Intonationsphrasierung, d. h. von längeren Phrasen, ausgegangen werden. Es werden theoretische Aspekte und frühere

Forschungsergebnisse zu den einzelnen prosodischen Merkmalen in den jeweiligen Interimsprachen behandelt, welche hauptsächlich in der produktiven Prosodie verortet sind. Danach werden sowohl Problemfelder und Haupttrends als auch Forschungslücken in der Prosodie im Zweitsprachenerwerb identifiziert und zusammengefasst. Auf der Basis vorheriger Studien folgt schließlich die formale und funktionale Darstellung der vier ausgewählten prosodischen Merkmale in der englischen Muttersprache und den drei Nichtmuttersprachlergruppen, woraus schlussendlich die Vorgehensweisen und Hypothesen[89] für die Untersuchung abgeleitet werden. Die prosodischen Merkmale sind folgende: Töne (Akzenttöne, Phrasentöne und Grenztöne) und Tonhöhenverläufe/Melodien (Akzent-, Phrasen- und Grenztonkonfigurationen), Intonationsphrasierung (Intonationsphrasen und Intermediärphrasen), Grundfrequenzumfang (*f0 level* und *span*) und *Uptalk*.

Zusammenfassend kann beobachtet werden, dass die frühere Forschung im Bereich der Prosodie bei Englischlernern mit anderen Muttersprachen ähnliche Defizite in der Frequenz, Variation und Realisierung von prosodischen Merkmalen in den verschiedenen Interimsprachen aufgedeckt hat.

Das allgemeine Bild, das sich aus den früheren Studien ergibt, ist, dass sich die Prosodie der Muttersprache erheblich von der der Lernersprache unterscheidet. Nichtmuttersprachliche Äußerungen sind durch ein reduziertes Toninventar, kürzere IUs (*intonation units*, dt. Intonationseinheiten), häufiger unterbrochene IUs, eine geringere phonetische Ausdehnung tiefer Töne und somit einen engeren Grundfrequenzumfang gekennzeichnet. Die Ergebnisse bezüglich der Variabilität zwischen den Sprechern und der Positionsempfindlichkeit der Grundfrequenz haben jedoch Gegenbeweise für einen engeren Grundfrequenzumfang erbracht, wodurch die muttersprachlichen Werte angenähert werden. Frühere Untersuchungen ergaben auch, dass die Sprachproduktion der Lerner in Bezug auf Sprachrhythmus, Artikulationsrate, Sprechgeschwindigkeit und durchschnittliche Äußerungslänge tatsächlich langsamer ausfällt. Schnelle Sprache führt selbst bei nichtmuttersprachlichem Output zu einer geringeren Anzahl sowie zu längeren IUs, während langsamere Sprache zu immer kürzeren und häufigeren IUs führt. Letztere haben zudem teils zur Folge, dass Nichtmuttersprachler des Englischen tendenziell auch vermehrt Akzenttöne verwenden. Insbesondere deutschen und spanischen Lernern wurde in der Vergangenheit attestiert, sie erzeugen mehr mittlere Töne bzw. ersetzen hohe durch tiefe Töne (und umgekehrt) in Kontexten in denen das jeweilige Gegenteil angemessener wäre. Es wurde auch dokumentiert, dass Nichtmuttersprachler des Englisch dazu neigen, ihre IPs mit einem

[89] Da es sich um 26 individuelle Hypothesen handelt, werden diese hier aus Platzgründen nicht aufgelistet.

höheren oder einem mittleren Ton zu beenden und dass Englischlerner tiefe Töne erzeugen, bevor hohe und komplexe Töne später im Zweitsprachenerwerbsprozess erlernt und produziert werden. Die meisten Abweichungen wurden durch eine vermeintliche Unsicherheit oder mäßige Angst seitens der Lerner beim Sprechen in ihrer L2 erklärt. In der vorliegenden Studie beabsichtige ich unter anderem, Uptalk in Lernersprache genauer zu erläutern. Dabei argumentiere ich, dass die hohen Töne der Lerner am Ende nicht nur auf Unsicherheiten beruhen, sondern auch für unterschiedliche pragmatische Funktionen eingesetzt werden. Frühere Studien zur L2-Prosodie haben gezeigt, dass die Lerner bei der Produktion der Prosodie abweichen. Welche Faktoren diese Abweichung bestimmen, ist jedoch noch nicht final geklärt (vgl. Mennen und De Leeuw 2014: 187).

Insgesamt hat sich gezeigt, dass Prosodie stark kontextabhängig ist und verschiedene Funktionen erfüllt. Die Variation wurde analysiert, um festzustellen, wie sie unter anderem mit Kontextfaktoren wie dem Geschlecht des Sprechers oder dessen Alter korreliert bzw. mit den unterschiedlichen Sprechhandlungen (Fragen, Antworten, Aussagen usw.) und der Sprechposition (mediale und finale Position in IUs). Jedoch fokussierte sich die bisherige Forschung vornehmlich auf muttersprachliches Englisch, während akustische Analysen zur Identifikation von Unterschieden bezüglich der Form und der pragmatischen Funktion in Lernerkorpora bisher die Ausnahme darstellen (z. B. Ramírez Verdugo 2005, 2006b; Ritchart und Arvaniti 2014). Dennoch misst aktuelle Intonationsforschung der pragmatischen Funktion von prosodischen Realisierungen eine signifikante Bedeutung bei (siehe auch GER), da diese für die Perzeption des Lerneroutputs eine entscheidende Rolle spielt. Pragmatik wird oftmals von Zuhörern nicht mit dem ersten Satz wahrgenommen, wirkt aber umso stärker, wenn von der muttersprachlichen Norm abgewichen wird. Da die Intonation für die Gewichtung des Wahrheitsgehalts einer Äußerung von immenser Bedeutung ist, sind pragmatische Fehler auf Grund unangemessener Intonationsmuster als gravierender anzusehen, als beispielsweise lexikalische Unstimmigkeiten. Zur pragmatischen Kompetenz zählt zum Beispiel auch die Kenntnis darüber, wie verschiedene Sprechakte prosodisch realisiert werden. Bei Lernern sind diese Fertigkeiten jedoch oftmals nur in geringem Maße vorhanden. Gleichermaßen sind sie sich über deren Effekte nicht bewusst.

Was Themen wie Uptalk und die detaillierte Analyse von Intonationsphrasen in der Sprachproduktion deutscher, tschechischer und spanischer Lerner anbelangt, so sind diese praktisch unerforscht. Darüber hinaus enthalten die meisten Studien zur L2-Prosodie nur sehr wenige Teilnehmer, analysieren jeweils nur ein Geschlecht, konzentrieren sich jeweils auf eine extralinguistische Variable, analysieren nur ein prosodisches Merkmal isoliert und / oder analysieren meist gelesene Sprache (*read speech*). Insbesondere ältere Studien untersuchen nur die Sprache von Lernern im Vergleich zu einem Modell-Muttersprachler (meist ein Dozent oder ein sehr kompetenter Redner) (z. B. Gutiérrez Díez 2008), obwohl gezeigt wurde, dass es innerhalb der Sprachen/Dialekte und Sprecher eine beträchtliche Variabilität gibt. Darüber hinaus wird natürlich vorkommende Sprache selten im Hinblick auf Prosodie analysiert (z. B. Tyler 2019).

Das vorliegende Dissertationsprojekt stellt im Folgenden eine Analyse vor, die diese vorherigen Ansätze und Forschungsergebnisse zu den vier ausgewählten prosodischen Merkmalen zusammen führt, miteinander verbindet und kombiniert, um so einen umfassenderen Einblick in die produktive Prosodie fortgeschrittener Lernergruppen mit unterschiedlichen Muttersprachen aus verschiedenen Sprachfamilien zu gewinnen. Dabei möchte ich nicht nur L2-prosodische Muster für die deutschen, tschechischen und spanischen Sprecher des Englischen beschreiben und vorhersagen, sondern auch wertvolle Forschungsergebnisse liefern, die die von Mennen (2015) aufgestellten LILt-Hypothesen untermauern.

Kapitel 3

Datenbasis und Methodologie

Als Datenbasis für die vorliegende Dissertation dienten die deutsche, tschechische und spanische Komponente des LINDSEI-Korpus. LINDSEI (*Louvain International Database of Spoken English Interlanguage*) ist ein internationales Projekt, das von der Université catholique de Louvain-la-Neuve (Belgien) initiiert wurde und beinhaltet gesprochene Lernerdaten mit verschiedenen L1-Hintergründen. Die ausgewählten Lernernkorpora beinhalten jeweils 50 Aufnahmen pro Korpus, welche orthographisch transkribiert und *error*-getaggt wurden (vgl. Kämmerer 2009). Die ausgewählten Komponenten wurden jeweils in unterschiedlichen Universitäten aufgenommen: Während die tschechischen Aufnahmen an der Karlsuniversität in Prag entstanden sind (LINDSEI-CZ; Gráf 2017) und LINDSEI-GE an der Justus-Liebig-Universität in Gießen kompiliert wurde (Brand and Kämmerer 2006), hat die spanische Komponente ihren Ursprung an zwei verschiedenen Universitäten in Spanien – Madrid und Murcia (Pérez-Paredes 2010).

Zum Vergleich wurden drei muttersprachliche Korpora herangezogen: 1) das an der Lancaster Universität in Großbritannien erstellte *Louvain Corpus of Native English Conversation* (LOCNEC; cf. De Cock 2004), 2) Die *New South Voices* (NSV) Sammlung, die hauptsächlich amerikanische Sprecher aus den Südstaaten der USA enthält, und 3) das

Nationwide Speech Project Korpus (NWSP; Clopper und Pisoni 2006), welches amerikanische Muttersprachler aus dem Norden, Westen, Mittelland und Süden der USA beinhaltet. Das LOCNEC Korpus dient dem direkten Vergleich mit den drei Lernerkorpora, da diese nach dem gleichen Korpusdesign aufgebaut wurden. Die Aufnahmen enthalten demnach alle drei Teile: einen Monolog zu drei möglichen Themen (Film, Land oder Erfahrung), ein informelles Interview und eine Bildbeschreibung.

Im autosegmentalen Rahmen berichtet diese Studie über eine statistische Analyse und qualitative Interpretation von Intonationsabweichungen von L2-Lernern in spontaner monologischer und dialogischer Sprache, die aus einer "Contrastive Interlanguage Analysis" (CIA) und einer "Contrastive Analysis" (CA) (vgl. Granger 1996, 2015) der deutschen, tschechischen und spanischen Sprache abgeleitet wurde. Hauptsächlich verfolgte die vorliegende Arbeit einen quantitativ-statistischen Ansatz, beinhaltet aber auch einen impliziten qualitativen Anteil. Die gemischte Methodik umfasst dabei sowohl korpusbasierte, korpusgesteuerte, instrumentelle, auditorische, und quantitative Methoden, als auch multivariate Analysen zur Untersuchung muttersprachlicher und nicht-muttersprachlicher Daten. Um die Analyse der prosodischen Merkmale zu ermöglichen, mussten alle Daten prosodisch annotiert werden. Dafür wurde, wie bereits erwähnt, das *Tone and Break Indices* (ToBI) (Silverman et al. 1992) System ausgewählt und mit *Praat* (Boersma and Weenink 2019) annotiert. ToBI wurde als prosodisches Annotationssystem ausgewählt, da es erwiesenermaßen für englische Intonationsstudien am häufigsten verwendet wird und in der Vergangenheit wiederholt auch für unterschiedliche Varietäten des Englischen sowie für andere Sprachen genutzt wurde. Einen weiteren, zentralen Vorteil stellt die klare Trennung phonetischer und phonologischer Level dar.

Die ausgewählte Korpusstichprobe in dem vorliegenden Dissertationsprojekt besteht aus der Sprachproduktion von insgesamt 225 Sprechern, die sich in Alter, Geschlecht, ihrer L1, ihren Sprachkenntnissen, der Länge der Auslandsaufenthalte im englischsprachigen Raum sowie in ihrer Interviewsituation unterscheiden. Um die Vergleichbarkeit der Daten zu gewährleisten, enthält die Korpusstichprobe ähnliche Sprachaufnahmen zu ähnlichen Themen, in einer dyadischen Interviewsituation, einer ähnlichen Altersgruppe (18-33, 22 Jahre im Durchschnitt), und der gleichen Sprechdauer (≈560 Wörter, 1,9 – 5,9 Minuten, 3,13 Minuten im Durchschnitt), die nur aus deklarativen Sätzen der Sprecher besteht. Somit eignen sich die ausgewählten Daten hervorragend für einen kontrastiven Vergleich und eine statistische Analyse. Für die Analyse der Dateien wurden die prosodischen Merkmale automatisch extrahiert und Signifikanztests mit Hilfe von Praat-Skripten (Boersma und

Weenink 2019) und R-Skripten (R Development Core Team 2015) durchgeführt. Die vorliegende Studie wählt einen neuen methodischen Ansatz, da verschiedene extralinguistische Variablen und komplexe statistische Modelle einbezogen werden.

Kapitel 4

Datenanalyse der prosodischen Merkmale

In Kapitel 4 wird über die Ergebnisse der quantitativen Studie berichtet. Hierbei wird rein frequenzbasiert auf die signifikanten Abweichungen der Lerner von der Muttersprachlernorm in den vier ausgewählten prosodischen Merkmalen eingegangen. Für die Untersuchung dieser wurden unterschiedliche lineare Regressionsmodelle erstellt. Jedes prosodische Merkmal (einschließlich der verschiedenen Maßeinheiten) wurde sowohl aus einer beschreibenden als auch aus einer inferenziellen Perspektive analysiert (soweit die verfügbaren Daten dies zuließen). Nach der Präsentation der deskriptiven Ergebnisse (hauptsächlich anhand von *ggplots*) wurde jede abhängige Variable in ein statistisches Modell eingegeben, wobei jedes der Modelle auf Grundlage der abhängigen Variable ausgewählt wurde, d.h. logistische Regressionen für numerische Werte mit nur einem Datenpunkt pro Sprecher, Modelle mit gemischten Effekten für mehrere numerische Werte pro Sprecher und binomiale logistische Regressionen für kategorische Daten mit zwei Ebenen. Die Ergebnisse wurden mit Hilfe von Tabellen und Effektdiagrammen dargestellt. Die britischen Muttersprachler (LOCNEC) wurden als *Intercept* (dt. Absolutglied) für die meisten Modelle verwendet, bei denen ein Vergleich zwischen Muttersprachlern und Lernern durchgeführt wurde. In einem zweiten Schritt wurden die tschechischen Fremdsprachenlerner als Intercept ausgewählt, wenn nur die Lerner miteinander verglichen wurden. Tabelle 1 fasst die Ergebnisse der prosodischen Merkmale zusammen. Insgesamt wurden 17 unterschiedliche Maßnahmen (*measures*) ausgewählt, die zeigen, wie sich die Lerner in Bezug auf diese prosodischen Merkmale von der Muttersprachlernorm unterscheiden. Zwar bewegen sich die Abweichungen aller Lerner von der Muttersprachlernorm in die gleiche Richtung, die Stärke der Abweichungen variiert jedoch. Wie aus der Tabelle ersichtlich ist, wurden diese entsprechend kategorisiert und erstrecken sich von dem stärksten Grad der Abweichung (1) bis hin zur geringeren Abweichung (4), welche aber trotzdem noch signifikante Unterschiede zur Muttersprachlernorm aufweist. Wenn Gruppen einen gleichen Wert für eine Variable erzielen, dann erhalten sie die gleiche Zuordnung. In den grauen Feldern in Tabelle 1 sind prosodische Merkmale hervorgehoben, die sich in unterschiedliche Richtungen der Divergenz bewegen.

Leere Felder bedeuten, dass sich die Lernergruppen für die betreffenden prosodischen Merkmale nicht signifikant von der Muttersprachlernorm unterscheiden.

Lediglich für die Häufigkeit der IUs konnte für deutsche und tschechische Lerner eine Abweichung von diesem Trend beobachtet werden, ebenso wie für Lerner aus Murcia, bei denen die Akzenttonhäufigkeit eine variierende Ausprägung aufweist. Während es also Unterschiede in der Verteilung von Akzenttönen und IUs für ein paar der L1-Gruppen gibt, zeigen die Lerner ähnliche Abweichungen in allen anderen Verteilungen, Realisierungen und der pragmatischen Funktion der untersuchten prosodischen Merkmale. Die größten Unterschiede zu den Muttersprachlern konnten in Bezug auf die Lerner aus Murcia gemessen werden, die wiederholt die extremsten Werte – insbesondere bei der Verteilung der Akzenttöne und dem Grundfrequenzumfang – erzielten. Die deutschen Lerner divergieren ähnlich häufig von den Muttersprachlern, produzieren jedoch die geringsten Abweichungen (z.B. bei der Verteilung von Akzenttönen und Intonationsphrasen). Die Lerner aus Madrid und Murcia ähneln sich sehr stark, wobei die Lerner aus Madrid jedoch insgesamt bessere Werte erreichen. Interessanterweise lässt sich - in Bezug auf die Stärke der Abweichung der Sprechgeschwindigkeit von IUs – eine Ähnlichkeiten zu den tschechischen Lernern beobachten.

Tabelle 1. Übersicht aller nicht-muttersprachlichen prosodischen Merkmale in den vier Interimsprachen

	Prosodische Merkmale	**Tschechische Lerner aus Prag**	**Deutsche Lerner aus Gießen**	**Spanische Lerner aus Madrid**	**Spanische Lerner aus Murcia**
Intonations-einheiten	**IPs**	1) weniger IPs	1) häufigere IPs	---	---
	ips	1) weniger ips	---	---	---
	IP-Sprech-geschwindig-keit (SG)	2) langsamere SG	3) langsamere SG	1) langsamste SG	3) langsamere SG
	ip-SG	1) langsamste SG	3) langsamere SG	1) langsamste SG	4) langsamere SG
Töne	**Tonhäufig-keit**	---	---	---	---
	Phrasen- und Grenzton-häufigkeit	2) häufigere hohe PT und GT	1) meisten hohen PT und GT	3) häufigere hohe PT und GT	4) häufigere hohe PT und GT
	Akzentton-häufigkeit	2) häufigere tiefe AT	1) meisten tiefen AT	3) häufigere tiefe AT	1) häufigere hohe AT
Häufigkeit von pragmatische Funktionen		1) variieren am meisten	2) zahlreiche unterschiedliche	3) geringfügig mehr	4) geringfügig mehr
	***F0 span* (gemessen in Halbtönen) HL-Melodien**	3) enger	2) enger	1) engste	1) engste
	***F0 span* LH-Melodien**	3) breiter	2) breiter	1) breiter	1) breiter
	F0 span	4) breiter	3) breiter	2) breiter	1) breitester

	HH-Melodien				
	***F0 span* für LL-Melodien**	---	---	---	1) breitester
Grund-frequenz-umfang	***F0 level* (gemessen in Hertz) HL-Melodien**	---	---	---	---
	***F0 level* für LH-Melodien**	2) höher	2) höher	2) höher	1) höchste
	***F0 level* für HH-Melodien**	1) höher	2) höher	1) höher	2) höher
	***F0 level* für LL-Melodien**	---	---	---	1) höchste
	Annäherung der Grund-frequenz an die des Interviewers (*entrainment*)	---	---	1) nähern sich weniger an	1) nähern sich weniger an
Anzahl abweichender prosodischer Merkmale insgesamt		12/17 (71%)	11/17 (65%)	11/17 (65%)	13/17 (77%)
Anzahl extremer prosodischer Merkmale		5/17 (29%)	3/17 (18%)	6/17 (35%)	8/17 (47%)

Die deutschen und tschechischen Lerner ähneln sich am meisten in Bezug auf die Verteilung der Töne (Akzent-, Phrasen-, und Grenztönen) und die Variation unterschiedlicher pragmatischer Funktionen. Während die deutschen Lerner am meisten in der Verteilung von Tönen abweichen, sind es für die tschechischen Lerner die pragmatischen Funktionen, die die größten Divergenzen verursachen. Die Sprache der deutschen und tschechischen Lerner war durch eine hohe Uptalk-Frequenz gekennzeichnet, für die keine signifikanten phonologischen Unterschiede zwischen den beiden Lernersprachen festgestellt werden konnten. Aus phonetischer Sicht erzeugten die deutschen und tschechischen Lerner in den Uptalk-Melodien eine sehr ähnliche Grundfrequenz.

Auch in Bezug auf andere Variablen zeigt sich eine klare Varianz zwischen den muttersprachlichen Daten und denen der Lernerpopulation, wobei sich die Lernergruppen auch weiterhin in eine ähnliche Richtung bewegen. So ist die Sprechgeschwindigkeit in Wörtern pro Sekunde (WPS) pro IU signifikant langsamer und die durchschnittliche Länge der IUs in Sekunden ist tendenziell länger. Es werden vermehrt hohe Phrasen- und Grenztöne verwendet sowie ein höherer und breiterer Grundfrequenzumfang für hohe Tonverläufe (HH- und LH-Melodien). Des Weiteren wird eine engere Grundfrequenzspanne (*f0 span*) für hoch-tiefe (HL) Tonverläufe produziert. Die meisten Lerner verwenden eine pragmatisch vielfältigere Prosodie (z.B. Betonung, Unsicherheit, Verständnisüberprüfung usw.), welche von einem höheren Bedürfnis nach Verständlichkeit motiviert zu sein scheint.

Ein sehr ähnliches Verhalten der Fremdsprachenlerner und Muttersprachler konnte auf der pragmatischen Ebene beobachtet werden. Finalität sowie Sprecherwechsel (*handing over of turns*) werden in beiden Gruppen durch tiefe Töne, Kontinuität und Verständnisüberprüfung (*checking*) hingegen durch hohe Töne signalisiert, wobei die meisten Lerner hierbei eine verstärkte Realisierung beobachten lassen, indem sie eine tiefere Grundfrequenz für Finalität und die Übergabe des Rederechts produzieren und eine höhere für Kontinuität und Verständnisüberprüfungen. Insgesamt verwenden die Lerner zur Markierung von Kontinuität häufiger hohe Phrasen- und Grenztöne am Ende von IUs. Sie sind demnach in den meisten Fällen in der Lage ihre Tonhöhenverläufe zu manipulieren, um ihr Rederecht aufrechtzuerhalten oder abzugeben. Dass Lerner ihr Rederecht verlieren, wurde selten beobachtet. Dies passiert zumeist, wenn sich *hesitation phenomena*, also jene Phänomene, die seitens des Sprechers eine Verzögerung herbeiführen, häufen, bzw. wenn vermehrt tiefere Töne verwendet werden. Analysen der Stimmqualität (*creaky voice*, dt. Knarrstimme), und *fluency phenomena*, also Flüssigkeitsphänomene, (gefüllte und ungefüllte Pausen, Häsitationen, Dehnungen von Silben) zeigen weiterhin Gemeinsamkeiten und Abweichungen von der Muttersprachlernorm. Während sich innerhalb der Muttersprachlerdaten (AmE und BrE) ein häufiger Gebrauch der Knarrstimme zeigt, die dazu beitragen kann, IPs voneinander abzugrenzen, verwenden die Lerner unterschiedliche Strategien und Flüssigkeitsphänomene, die einen höchst signifikanten Einfluss auf die Häufigkeit von Intonationsgrenzen haben. Diese unterscheiden sich je nach L1-Gruppe und zeigen weiterhin sprecherabhängige Präferenzen. Die Äußerungen der Sprecher werden in Abschnitte unterteilt, welche durch eigene Grundfrequenzverläufe gekennzeichnet sind und werden je nach L1-Gruppe durch unterschiedliche Strategien, d.h. kurze gefüllte oder ungefüllte Sprechpausen, Atempausen oder eine Verlängerung/Dehnung von Tonbewegungen und / oder Silben realisiert. Die tschechischen Sprecher des Englischen erzeugen eine sehr fließende Sprache, zeigen jedoch gleichzeitig eine übermäßige Verwendung von gefüllter Pausen sowie verlängerter Silben, welche an IPs grenzen. Dabei produzieren die tschechischen Sprecher hohe Phrasen- und Grenztöne, die die Äußerungen miteinander verknüpfen und nahezu endlose IUs erzeugen. Die deutschen Lerner machen von einer anderen Strategie zur Intonationsphrasierung Gebrauch. Sie verwenden anhaltende, extrem hohe Phrasen- und Grenztöne in Verbindung mit ungefüllten Pausen, was schlussendlich zur Überproduktion von IUs führt. Die Ergebnisse bestätigen und erweitern frühere Studien und deuten sehr stark darauf hin, dass ein systematisch abweichendes und universales Interimsprachensystem existiert. Die einzigen muttersprachlich ähnlichen prosodischen Merkmale sind die Tonfrequenz, die Tatsache, dass Lerner sich – wie auch die Muttersprachler – dem Grundfrequenzumfang der Interviewer anpassen (*f0 entrainment*), dass monotonale Akzenttöne bitonalen bevorzugt werden, die Höhe der Grundfrequenz (*f0 level*) von HL-Melodien und die Häufigkeit von Intonationsgrenzen – wobei Letzteres nur für spanische Lerner des Englischen zutrifft.

Um die Ergebnisse besser zu erklären, wurde eine multifaktorielle Regressionsanalyse durchgeführt, die folgende Faktoren beinhaltet: Sprechstil, Alter, Geschlecht, Auslandsaufenthalte, pragmatische Funktionen (Fortsetzung/Kontinuität, Auflistung usw.), Sprechakte (Aussagen vs. Antworten) usw. Die prosodischen Merkmale und der Einfluss der untersuchten Prädiktoren sind in Tabelle 2 zusammenfassend aufgeführt. Die Haupteffekte, die die Variation der L2-Prosodie begründen, liegen im Geschlecht der Sprecher und in bestimmten Flüssigkeitsvariablen, während lediglich geringfügige Auswirkungen auf das GER-Niveaus, das Alter, den Sprechstil und den Auslandsaufenthalt zurückgeführt werden können. In Bezug auf den Sprechstil wurde schon in früheren empirischen Studien gezeigt, dass Lerner nur sehr selten Zugang zu verschiedenen Registern und den damit verbundenen Unterschieden haben und daher die Varianz zwischen monologischen und dialogischen Sprechstilen in den Lernerdaten keine große Rolle spielt. Abschließend kann gesagt werden, dass alle analysierten Prädiktoren zusammen zur Form der L2-Prosodie der drei untersuchten Lerngruppen beitragen. In Bezug auf die theoretischen Konzeptualisierungen der L2-Prosodie deuten die Ergebnisse darauf hin, dass die Einbeziehung lediglich eines Kompetenzmaßes möglicherweise nicht für alle prosodischen Merkmale ein vollständiges Bild ergibt. Jeder Prädiktor beleuchtet ein anderes prosodisches Merkmal und eine andere Intonationsdimension. Dies wird in der Diskussion weiter erörtert, in der ein multifaktorielles Modell vorgeschlagen wird, welches versucht, die Ergebnisse der vorliegenden Studie zu erläutern. Dies hat insbesondere Auswirkungen auf Sprachtests und -bewertungen sowie auf die GER-Deskriptorskalen. Insgesamt wurden einige der Annahmen des LILt-Modells, die auf Grundlage des vorliegenden Datensatzes testbar waren, durch die Korpusanalyse untermauert.

Tabelle 2. Übersicht der Zweitsprachenerwerbsprädiktoren und deren Einfluss auf die prosodischen Merkmale[90]

	Alter	**Geschlecht**	**Auslands-aufent-halt**	**GER-Niveau**	**Sprechstil**	**Flüssigkeit**	**Interviewer Geschlecht**	**Anzahl weiterer Fremd-sprachen**	**Anzahl der Jahre des Englisch-sprach-erwerbs an der Universität**	**Anzahl der Jahre des Englisch-sprach-erwerbs an der Schule**
Ton-häufigkeit	X	X			X	X				
Hohe Phrasen- und Grenztöne		X	X		X		X	X		X
IP Häufigkeit		X			X	X				
IP Sprech-geschwindig-keit				X	X	X		X		
ip Häufigkeit		X				X			X	
ip Sprech-geschwindig-keit	X			X		X				
F0 level		X			X	X				
F0 span		X	X		X	X				
***Uptalk* Häufigkeit**	X	X	X		X					

[90] X bedeutet, dass ein Prädiktor einen signifikanten Einfluss auf das entsprechende Merkmal hat, die Richtung des Einflusses wird jedoch nicht angezeigt. Diese wird in Kapitel 5, Abbildung 1 dargestellt und diskutiert.

Kapitel 5

Diskussion, Implikationen und das Erklärungsmodell

Es lässt sich zusammenfassend feststellen, dass sich die drei Lernergruppen in drei von vier Intonationsdimensionen von der Muttersprachlernorm unterscheiden, wobei der Unterschied vorrangig in der Stärke und nicht der Richtung der Abweichung deutlich wird. Die Verteilungs-, Funktions- und Realisierungsunterschiede können auf die Muttersprache, das Geschlecht, die Rolle und das Geschlecht des Interviewers, den Sprechstil, Flüssigkeitsvariablen, das Sprachniveau und positionsempfindliche Unterschiede (d. h. Position der Töne innerhalb einer IU) zurückgeführt werden. Der Zusammenhang dieser Variablen und deren Einfluss auf die untersuchten prosodischen Variablen werden in Abbildung 1 und den darauffolgenden Absätzen erläutert. Die vorliegende Arbeit resultierte in der Entwicklung eines Modells, das die Rolle von unterschiedlichen (extra-)linguistischen, kontextuellen, und intonatorischen Faktoren für die prosodischen Realisierungen von Lernern zum Gegenstand hat. Mennens (2015) LILt Modell stellt den Ausgangspunkt für das erstellte Modell dar, das sich aus den Resultaten der empirischen Untersuchung ableitet. Wie im Folgenden ersichtlich wird, geht die graphische Darstellung auf die Richtung und Stärke der Faktoren – hier durch die Dicke der Pfeile kenntlich gemacht – und deren Auswirkungen auf die Intonationsdimensionen ein.

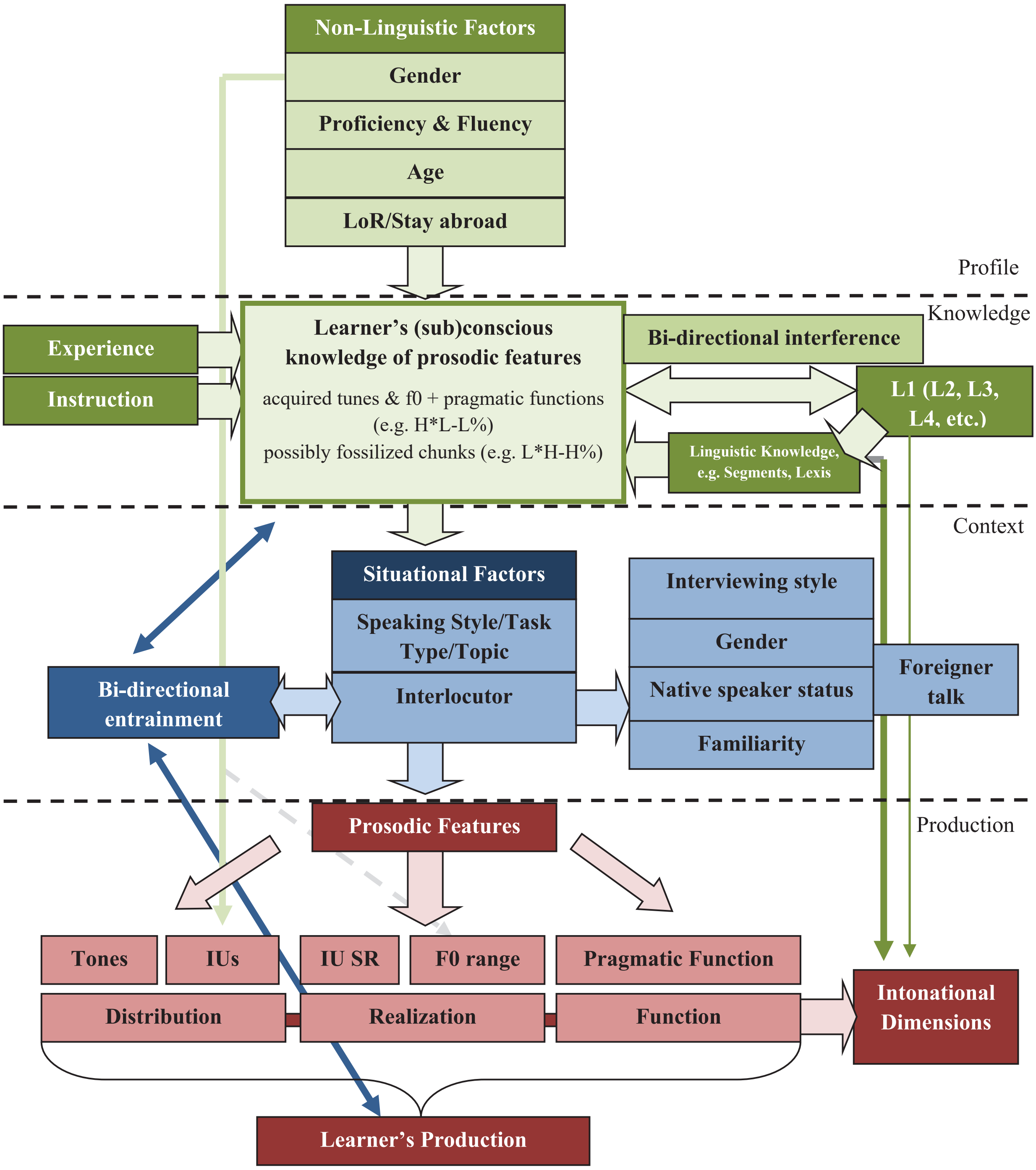


Abbildung 1. Erklärungsmodell der prosodischen Produktionen der deutschen, tschechischen und spanischen Lerngruppen. SR = Speech Rate, IU = Intonation Unit

Zusammenfassend konnten vier miteinander verbundene Ebenen identifiziert werden, die die finale prosodische Produktion der Lerner beeinflussen. Die erste Ebene umfasst das Profil der Lerner, das unterschiedliche extralinguistische Variablen einschließt, welche im Modell nach ihrer Wichtigkeit in der vorliegenden Studie sortiert wurden. Das Geschlecht der Lerner hatte den größten Einfluss auf die Verteilungsebene der Intonationsdimensionen, da große geschlechterspezifische Unterschiede in der Verteilung von IUs beobachtet werden konnten.

Natürlich hat das Geschlecht auch einen großen Einfluss auf die Realisierungsebene im Grundfrequenzumfang, welcher jedoch aus biologischen Gründen gegeben ist, während die Verteilungsunterschiede von IUs eher auf soziale und diskursspezifische Unterschiede zwischen den Geschlechtern hinzuweisen scheinen. Der Lerner kann nur prosodische Merkmale erzeugen, die innerhalb seiner physiologischen Grenzen (Geschlecht, Alter) und seines erlernten Verhaltens (sozial akzeptierte Normen des Grundfrequenzumfangs für ein bestimmtes Geschlecht) liegen. Während die Flüssigkeit der Sprache, die ich als Teil des Sprachniveaus der Lernerprofile erachte, einen starken Einfluss auf fast alle prosodischen Merkmale hat (siehe Tabelle 2), haben das Alter und der Auslandsaufenthalt der Lerner einen eher nebensächlichen Effekt.

Die zweite Ebene befasst sich mit dem Wissen des Lerners, welches (unter-)bewusst erworbene Melodien und Grundfrequenzen (mit oder ohne deren pragmatische Funktionen) sowie möglicherweise fossilisierte Einheiten, d.h. solche, die sich über einen gewissen Zeitraum gefestigt haben, zusammen mit deren Prosodie umfassen kann. Dieses Wissen wird weiter beeinflusst durch den Kontakt des Lerners mit Muttersprachlern und der Sprache im Allgemeinen, den (prosodischen) Sprachunterricht sowie die Muttersprache der Lerner. Letztere kann auch bidirektional beeinflusst, d.h. die prosodischen Englischkenntnisse des Lerners können auf die L1 übertragen werden. Weitere Fremdsprachen (L3, L4 usw.), die ein Lerner spricht oder kennt, sind ebenfalls im Modell enthalten, da die Ergebnisse zeigen, dass eine zunehmende Anzahl weiterer Fremdsprachen zu einer muttersprachlichen Häufigkeit von hohen Tönen und einer entsprechenden IP-Sprechgeschwindigkeit führt. Die linguistische Kompetenz der Lerner hat auch einen Einfluss auf die Auswahl der erlernten prosodischen Muster, die weitere linguistische Ebenen – wie das Beherrschen von Segmenten (z.B. Dehnung von Segmenten oder Silben) und den Wortschatz der L2 – einschließen. Sie wird wiederum von der L1 beeinflusst und hängt vom Sprachniveau der Lerner ab. Es konnte festgestellt werden, dass die prosodischen Abweichungen eindeutig durch L1-Muster motiviert sein können, weshalb das Sprachniveau der Lerner als ein wichtigerer Faktor eingestuft wurde.

Die spanischen Lerner wurden beispielsweise als weniger kompetent im L2-Gebrauch eingeschätzt, was erklären könnte, warum die fortgeschritteneren deutschen und tschechischen Lerner insbesondere im Gebrauch von hohen Phrasen- und Grenztönen mehr Gemeinsamkeiten aufzeigen. Es wurde vorgeschlagen, dass sich die fortgeschrittenen Lerner in einer *prosodischen Erkundungsphase* befinden, indem sie versuchen, sich der Muttersprachlernorm anzunähern. Im Falle der weniger kompetenten spanischen Lerner wurden Vermeidungsstrategien als Erklärung für den engsten Grundfrequenzumfang für HL-Melodien und den höchsten und breitesten für LH- und HH-Melodien vorgeschlagen. Möglicherweise sind die spanischen Sprecher mit dem lexikalischen Abruf beschäftigt und produzieren somit einen engen Grundfrequenzumfang für HL-Melodien (welche am häufigsten verwendet werden). Währenddessen deuten die LH- und HH-Melodien darauf hin, dass sie versuchen, ihre allgemein enge Grundfrequenz gelegentlich zu variieren und so weit wie möglich von ihrer Muttersprache abzuweichen, die, im Falle des Spanischen in Murcia, in der Forschung bereits als der engste Grundfrequenzumfang der spanischen Varietäten ermittelt werden konnte (Monroy and Hernández-Campoy 2015: 237).

Die dritte Ebene enthält den Kontext, in dem sich die Lerner befinden und somit die Situationsfaktoren, denen sie bei der Datenerhebung ausgesetzt waren. In der vorliegenden Studie wurden der Sprechstil / Aufgabentyp und das ausgewählte Thema analysiert, über das die Teilnehmer sprechen. Der Gesprächspartner übt den größten Einfluss auf die Produktionen des Lerners aus, da eine bidirektionale Annäherung der Grundfrequenz (*prosodic entrainment*) stattfinden kann, d.h. der Lerner ahmt den Gesprächspartner nach und umgekehrt. Die Intonationsmuster (insbesondere die *f0 range*) der Interviewer wirken sich nicht nur kurzfristig auf den finalen Output der Lerner aus, sondern können sich möglicherweise auch zu fossilisierten Intonationsmustern in deren Wissen manifestieren, welche durch einen regelmäßigen Kontakt mit *teacher/foreigner-talk* gefestigt werden können. Inwieweit der Gesprächspartner die prosodischen Produktionen des Lerners beeinflusst, hängt ferner vom Interviewstil, dem Geschlecht und dem Muttersprachlerstatus des Interviewers ab – bzw. davon, ob er oder sie *foreigner-talk* ausübt – sowie von dem Bekanntheitsgrad zwischen den Gesprächspartnern. Annäherungseffekte (*entrainment effects*) waren hauptsächlich auf der Ebene des Grundfrequenzumfangs, Unterschiede im Sprechstil hingegen auf der Intonationseinheitsebene sichtbar.

Schließlich zeigt die vierte Ebene, dass sich die verschiedenen prosodischen Merkmale, die ein fortgeschrittener Lerner (L1 Deutsch, Tschechisch oder Spanisch) hervorbringt, hauptsächlich in drei von vier Intonationsdimensionen unterscheiden können, d.h. in verteilenden, realisierenden und pragmatischen Funktionen (vgl. Ladd 2008; Mennen 2015). Dies bestätigt frühere Untersuchungen, die belegen konnten, dass die realisierenden und funktionalen Dimensionen der Intonation für die meisten Lerner Schwierigkeiten zu bereiten scheinen, während die systemische Dimension kein Problem darstellt (vgl. Mennen 2015; Ward und Gallardo 2017: 22). Die Pfeile zwischen den drei Dimensionen in Abbildung 1 zeigen an, dass sich die Intonationsdimensionen gegenseitig beeinflussen. Letztendlich wird die endgültige prosodische Produktion eines Lerners von allen drei Intonationsdimensionen beeinflusst, die je nach L1, linguistischer Kompetenz auf anderen linguistischen Ebenen und dem Geschlecht in der Stärke der Abweichungen variieren können.

Während sich die Lerner in nahezu allen Intonationsdimensionen des LILt-Modells und in allen untersuchten prosodischen Variablen unterschieden, ergaben die Analysen klare gruppenspezifische Trends: Während die spanischen Lerner am deutlichsten auf der Realisierungsebene von der Muttersprachlernorm abweichen (d.h. in Bezug auf die Sprechgeschwindigkeit pro IU und den Grundfrequenzumfang in den verschiedenen Melodien), weichen die tschechischen und deutschen Lerner am stärksten in der Verteilungs- und Funktionsebene durch die Auswahl der Tonkategorien ab, für die in beiden Gruppen der gleiche Trend beobachtet wurde. Auch die Abweichungen bezüglich der Vielfalt der für diese Tonkategorien ausgewählten pragmatischen Funktionen sowie die Häufigkeit der IUs zeigen entscheidende Unterschiede zu den muttersprachlichen Ergebnissen, wobei es möglich war, zwischen den beiden Lernergruppen ähnliche Trends bzw. – bezüglich der Frequenz der IUs – ein entgegengesetztes Verhalten festzustellen. Basierend auf diesen Ergebnissen wurden eher kompetenzbezogene (anstelle von L1-basierten) Erklärungen gegeben, da alle Lerner eine ähnlich abweichende Prosodie etabliert zu haben scheinen.

Die wiederholte Verwendung von Uptalk in der Lernersprache wurde als das auffälligste prosodische Merkmal identifiziert, ein Ergebnis, welches frühere Studien bestätigt (z. B. Gut 2009; Götz 2013). Zusammenfassend lässt sich festhalten, dass mit fremd klingendem (*foreign-sounding*) Uptalk einerseits ein Verteilungsproblem besteht, da es häufiger verwendet wird. Andererseits gibt es ein Realisierungsproblem, da kaum Variation in Form und Grundfrequenzumfang der hohen Töne und der Gesamtmelodie festgestellt werden konnte, sowie ein pragmatisches "Funktionsproblem", das auf die gleiche Form und tonale Parallelität zurückzuführen ist. Frühere Studien beobachteten einen häufigeren Gebrauch von hohen Phrasen- und Grenztönen, welche oftmals auf Unsicherheiten seitens der Lerner zurückgeführt wurden. Aus linguistischer Perspektive stellt sich also die Frage, wie Unsicherheit sich im Sprachsignal der Lerner manifestiert. In der vorliegenden Studie wurde Uptalk als polysemischer pragmatischer Marker in der Lernersprache identifiziert, der unterschiedliche Formen einnehmen und unterschiedliche Funktionen erfüllen kann. Es wurde versucht, zu zeigen, dass Uptalk für verschiedene pragmatische Zwecke verwendet wird (ähnlich denen, die in früheren Untersuchungen gefunden wurden), die zu einem besseren Verständnis der Verwendung von Uptalk in der Muttersprache beitragen sollten (Pierrehumbert und Hirschberg 1990; Brasilien 1997; Wichmann 2000; Wennerstrom 2001;

Pickering 2009; Levon 2016: 139; Burdin und Tyler 2018). Insbesondere wird Uptalk in nicht-muttersprachlichem Englisch verwendet, um das Wort zu halten (i.e. für Kontinuität), um zu überprüfen (i.e. für Bestätigungsanfragen, Überprüfung des Verständnisses usw.), um Unsicherheit zu signalisieren (i.e. indem indirekt um Feedback gebeten wird) und um Elemente aufzulisten. Diese Ergebnisse deuten darauf hin, dass Lerner einen gemeinsamen Entwicklungspfad einzuschlagen scheinen, indem sie Uptalk in Situationen verwenden, in denen Muttersprachler dies nicht tun würden. Innerhalb dieses Disserationsprojektes wurden viele Erklärungen für die größere Häufigkeit hoher Phrasen- und Grenztöne vorgeschlagen – vom Einfluss der L1 über einen Mangel an pragmatischer Genauigkeit bzw. an pragmatischem Wissen, bis hin zum Einfluss von Interviewern (d.h. übertriebenem *foreigner-* oder *teacher-talk*) und damit der Bereitstellung eines anderen prosodischen Modells für die Lerner, von dem sie ihre prosodischen Muster ableiten. Einer der Hauptpunkte der vorliegenden Studie war es demnach, darzustellen, warum hohe Phrasen- und Grenztöne nicht einfach als Signalisierung von Unsicherheit seitens der Lerner gesehen werden sollten. Dies muss auch im Fremdsprachenunterricht und der Sprachbewertung Berücksichtigung finden und wurde deshalb in den fremdsprachlichen Implikationen weiter diskutiert.

Auf Grund der Ergebnisse der quantitativen Analyse wurden in Kapitel 5 einige fremdsprachendidaktische Implikationen für die Sprachbewertung und den Fremdsprachenunterricht abgeleitet. Zunächst werden kurz die wichtigsten Punkte zur Sprachbewertung zusammengefasst: Bei einem Lerner auf C1-C2-Niveau sollten die Bewerter hauptsächlich auf die Verwendung der Grundfrequenz achten, insbesondere darauf, ob die pragmatische Absicht klar ist, die praktisch frei von einem L1-Einfluss sein sollte. Für B1-B2-Niveau sollte der Fokus der gleiche sein, wobei die Bewerter hier durchaus noch Transferphänomene aus der L1 erwarten und auch akzeptieren können. Basierend auf meiner Analyse sind die wichtigsten Merkmale, die bewertet, berücksichtigt oder abgefragt werden könnten, die folgenden:

1. Töne und Grundfrequenzumfang: Die Verwendung von hohen Tönen am Ende von Intonationseinheiten macht Lerner *per se* nicht weniger kompetent in der L2 (d.h. weniger muttersprachlich). Der Bewerter/Lehrer muss auf Grund seiner Realisierung, Verteilung und Funktionen, die diese hohen Töne erfüllen können, zwischen kompetent genutzten und fremdklingenden hohen Tönen unterscheiden. Genau wie in der englischen Muttersprache, kann der Uptalk in der nicht-muttersprachlichen Produktion unterschiedlich verteilt und von Männern und Frauen für unterschiedliche

pragmatische Funktionen verwendet werden. Weibliche Nichtmuttersprachler des Englischen organisieren – ebenso wie weibliche Muttersprachler (vgl. Levon 2016) – ihren Diskurs möglicherweise komplexer. Wenn der Bewerter verschiedene pragmatische Funktionen von hohen Tönen identifizieren kann, könnte ein höheres Sprachniveau des Lerners angenommen werden, da die Sprache möglicherweise fließend genug ist und das lexikalische Abrufen möglicherweise automatischer erfolgt. Die Lerner können es sich daher erlauben, ihren Grundfrequenzmustern mehr Aufmerksamkeit zu schenken und den Gesprächspartner in das Gespräch einzubeziehen, indem sie ihren Diskurs mithilfe von Uptalk strukturieren. Hohe Töne sollten in der Bewertung stärker berücksichtigt werden als fallende Töne, da hohe Töne erst zu einem späteren Zeitpunkt im SLA-Prozess erworben werden (vgl. Grosser 1997; Gut 2009). Auch der Gebrauch von einer engeren Grundfrequenz am Anfang von Intonationseinheiten und eine deutlich breitere Grundfrequenz am Ende von Intonationseinheiten können Indikatoren eines L1-Einflusses sein, welcher auf dem C2-Niveau nicht mehr akzeptabel sein sollte. Fragen, die bei der Bewertung der Form und Realisierung von Tönen in unterschiedlichen Sprechstilen möglicherweise hilfreich sein könnten, sind folgende:

- Wiederholen die Lerner die gleiche Form von Tönen (anhaltend, steigend usw.) bzw. gibt es eine übertriebene Realisierung von Tönen (extrem hohe oder enge Grundfrequenz), um bestimmte pragmatische Funktionen zu signalisieren?
- Wenn Lerner beispielsweise an einem Monolog beteiligt sind, variieren sie ihre Verwendung von Tönen, um die Fortsetzung zu signalisieren und Feedback vom Interviewer zu erhalten oder verwenden sie die gleiche Form und Realisierung von Tönen?
- Verwenden sie eine extrem enge Grundfrequenz, um eine Übergabe des Rederechts in Dialogen zu signalisieren?
- Verwenden sie eine extrem hohe Grundfrequenz in Monologen, um Kontinuität zu signalisieren und/oder um das Verständnis zu überprüfen?
- Hier kann der Bewerter auf sein eigenes Verhalten achten: Löst der Lerner während er spricht intensives *Back-Channeling* (d.h. Feedback/Rückmeldung) und/oder sogar eine Frage oder Aussage bei mir aus? Gebe ich im Allgemeinen häufig minimale Rückmeldungen und kann daher beim Lerner vermehrt hohe Töne/Uptalk hervorrufen? Spreche ich in einem übertriebenen

teacher/foreigner-talk, der dazu führen kann, dass der Befragte sich meiner eigenen Sprache anpasst (*entrainment*)? Sind meine eigenen prosodischen Produktionen Teil des phonetischen Repertoires des Lerners (basierend auf Geschlechterpaarung und gesprochenem/angestrebtem englischen Dialekt)?

  - Verwendet der Lerner auch pragmatische Marker oder verlässt er sich zu sehr auf die Intonation, um pragmatische Absichten zu signalisieren?

2. Intonationsphrasierung: Intonationsphrasen tragen auch zu einer effektiven und verständlichen Kommunikation bei, während ein übermäßiger Gebrauch von *fillers* und Tönen die Verarbeitung und das Verständnis von Äußerungen erschweren kann. Eine Verbindung von äußerungsbindenden und / oder trennenden Merkmalen zusammen mit der Verwendung von übermäßigen hohen und tiefen Tönen kann zu nicht-muttersprachlichen Intonationsphrasierungsmustern führen. Während auf dem C1-Niveau gelegentliche Fehler bezüglich prosodischer Merkmale gemäß GER zulässig sind, sollten diese auf dem C2-Niveau nicht mehr auftreten. Ebenfalls ist auf der Intonationsphrasierungsebene ein L1-Einfluss, der teilweise durch den Transfer des Rhythmus (z.B. Dehnung von Silben) und den von *fillers* aus der L1 bedingt ist, zu erwarten. Wenn diese Tendenzen einen reibungslosen und verständlichen gesprochenen Diskurs stören und der Bewerter Schwierigkeiten hat, zu entscheiden, wann er das Rederecht übernehmen oder wann er eine Frage stellen sollte, müsste der Lerner in Bezug auf seine Leistung negativer bewertet werden.

Das Unterrichten der in der vorliegenden Arbeit untersuchten prosodischen Merkmale sollte einerseits gezielt im Klassenzimmer stattfinden, andererseits jedoch auch an der Universität, wo Kurse angeboten werden sollten, die gezielt auf zukünftige Lehrer abgestimmt sind. Sehr häufig wird der Schulung der Prosodie ein niedrigerer Wert beigemessen als z.B. der Vermittlung von Grammatik. Da alle fortgeschrittenen Lerner in der vorliegenden Studie jedoch prosodische Abweichungen von der Muttersprachlernorm aufweisen, sollte dieser Aspekt von Lehrern in Schulen und insbesondere Dozenten an Universitäten aufgegriffen werden. Auf Grundlage der Ergebnisse, die belegen, dass alle Lerner ein ähnliches Verhalten zeigten, kann folgende Reihenfolge für die Instruktion vorgeschlagen werden: Die Häufigkeit von Tönen und das f0-Level von HL-Melodien scheint recht früh erworben zu werden und für die Lerner im fortgeschrittenen Stadium kein Problem darzustellen, da sie sich in diesen Messungen nicht von den Muttersprachlern unterscheiden. Die Tonwahl und die zugehörige Grundfrequenzspanne sowie die Sprechgeschwindigkeit von Intonationseinheiten scheinen den Lernern größere Probleme zu bereiten und sind noch nicht von allen Lernern erworben worden. Es gibt dennoch L1-basierte Unterschiede in Bezug auf den Erwerb der Intonationsgrenzenhäufigkeit, welche im Fremdsprachenunterricht Berücksichtigung finden sollte. Zusammenfassend können also auf Grundlage der Ergebnisse der vorliegenden Studie die folgenden Vorschläge für den Unterricht zukünftiger Lehrer gemacht werden:

- Zuerst sollten Lehrer mit den Studenten diskutieren, wie Prosodie definiert und analysiert werden kann. Aus Gründen der Konsistenz sollte das ToBI-Annotationssystem verwendet werden, anstatt die Studenten ihre eigenen prosodischen Symbole entwickeln zu lassen, was zu einer zusätzlichen Unsicherheit in Bezug auf die prosodische Annotation, Analyse und Terminologie führen könnte. Nicht nur die prosodischen Symbole, sondern auch die AM-Terminologie könnten verwendet werden, damit zukünftige Lehrer sich im Rahmen ihrer Ausbildung auch auf diese Begriffe beziehen und auf diese Weise zu einem allgemeineren Verständnis der prosodischen Merkmale beitragen können. Auch deren Schüler werden so mit der Terminologie vertraut gemacht und können diese im Unterricht verwenden.
- Lehrer sollten sich zuerst auf die Intonationsphrasierung und die Wirkung der Grundfrequenz auf die Kohäsion konzentrieren sowie auf das Diskursmanagement und den Beziehungsaufbau zum Gesprächspartner. Basierend auf der L1 der Lerner muss der Lehrer diese auf äußerungsbindende (oder -verknüpfende) und -trennende prosodische Muster und Flüssigkeitsphänomene aufmerksam machen. Dies unterstreicht erneut die Vernetzung von pragmatischem, flüssigkeitsbezogenem und prosodischem Wissen. Würden beispielsweise tschechische Lerner vermehrt Diskursmarker und weniger gefüllte Pausen oder Dehnungen am Ende von IUs verwenden, könnte dies dazu führen, dass ihre Intonationsphrasierung muttersprachlicher wird. Deutsche Lerner hingegen müssen auf die Verwendung von (extrem) hohen Tönen in Verbindung mit ungefüllten Pausen aufmerksam gemacht werden, die zur Überproduktion von Intonationseinheiten führen können. Spanische Lerner zeigten bezüglich der Intonationsphrasierung keine signifikanten Abweichungen von der Muttersprachlernorm, frühere Untersuchungen haben jedoch auch gezeigt, dass sie *fillers* von ihrer L1 auf die L2 übertragen und daher möglicherweise nicht-muttersprachliche Intonationsgrenzen erzeugen. Daher ist es wichtig, L1-

spezifische *filler*-Transferphänomene zusammen mit der Bildung von Intonationseinheiten zu betrachten und zu unterrichten.

- Prosodie sollte im Zusammenhang mit *pragmatic markers* zusammen mit Flüssigkeitsphänomenen unterrichtet werden. Hohe Töne und bestimmte Flüssigkeitsphänomene wurden in der vorliegenden Studie als äußerungsbindend aber auch –trennend identifiziert. Diese leisten somit einen wichtigen Beitrag zur Verbesserung der Flüssigkeitskompetenz. Für C2-Niveaus sollten sich die Lehrer auf die feineren Bedeutungsnuancen konzentrieren und darauf, wie Prosodie verwendet werden kann, um dieses Ziel zu erreichen. Insbesondere Töne und Melodien sollten von Lehrern und Studenten analysiert werden und auch wie diese eingesetzt werden können, um verschiedene pragmatische Funktionen – wie Kontinuität, Signalisieren neuer Informationen, Kohäsion und Abgrenzung, Verständnisprüfungen usw. – zu signalisieren. Liegt der Schwerpunkt auf der Pragmatik der Prosodie, so kann dies gleichzeitig die anderen Intonationsdimensionen verbessern. Wenn die Lerner beispielsweise wissen, wie sie hohe Töne angemessen verwenden, um Kohäsion zu schaffen, anstatt extrem hohe Phrasen- und Grenztöne zu verwenden, die als Unsicherheit aufgefasst werden können, wird gleichzeitig auch ihr Grundfrequenzumfang für diese Melodien enger und folglich muttersprachlicher.
- Lehrer sollten die Studenten auf die Angemessenheit der Verwendung von Uptalk aufmerksam machen und darauf, in welchen Situationen sie von Vorteil (z. B. im Klassenzimmer, siehe Pickering 2001) und wo sie eher nicht geeignet wäre (z. B. Vorstellungsgespräch).
- Lehrer könnten Studenten (zukünftige Lehrer in der Ausbildung) kleine Monologe vorbereiten lassen und andere Studenten bitten, diese anzuhören und “auffällige” prosodische Merkmale zu notieren. Diese Monologe könnten auch aufgenommen und genauer mit Hilfe von Sprachanalysesoftware analysiert werden. Lehrer sollten ihre Studenten auf ihre eigene Sprache aufmerksam machen und diese dazu veranlassen, über die Verwendung von hohen und tiefen Tönen als Kommunikationsmittel nachzudenken. Am besten wäre es, die Studenten in ein Interview einzubeziehen und ihnen die Aufnahmen vorzuspielen, um ihnen zu zeigen, wie sie hohe und tiefe Töne (und allgemeine prosodische Muster) im Gespräch selbst verwendet haben.

- Lehrer könnten ihre Studenten darum bitten, Aufzeichnungen eines muttersprachlichen Modellsprechers einer englischen Varietät zu sammeln, die sie als angenehm empfinden und / oder die sie selbst gerne sprechen würden. Dieser könnte beispielsweise an einem Dialog beteiligt sein. Mit Hilfe einer Sprachanalysesoftware (z. B. Praat oder Elan) könnten die Schüler die Grundfrequenz und die Intonationsphrasierung analysieren und Instanzen jeder pragmatischen Funktion finden und ihre prosodischen Realisierungen notieren (z. B. den für bestimmte pragmatische Funktionen verwendeten Grundfrequenzumfang). Dann könnten sie versuchen, einige der Äußerungen des Modellsprechers nachzuahmen und sie parallel mit ihren eigenen Produktionen zu vergleichen. Auf diese Weise hätte der Lehrer – neben einer Analyse der prosodischen Parameter – auch eine Sammlung pragmatischer Funktionen und könnte diese Sammlung für zukünftige Lernergruppen verwenden und erweitern. Ideal wäre es, eine öffentliche Plattform für Lehrer mit derartigen Daten zur Verfügung zu stellen. Wenn Korpusdaten verwendet werden, können diese natürlich nicht öffentlich zugänglich gemacht werden.
- Lehrer, insbesondere wenn sie selbst keine Muttersprachler sind, sollten sich mit verschiedenen Varietäten des Englischen sowie regionalen Unterschieden in Bezug auf die Prosodie auseinandersetzen (z. B. AmE vs. BrE) und den Studenten Ratschläge geben, wie sie sich mit den prosodischen Mustern einer bestimmten Varietät des Englischen vertraut machen können. Zukünftige Lehrer müssten dafür in speziellen Kursen an der Universität geschult werden.

Zuerst sollten Lehrer ein Verständnis für Prosodie und die dazugehörige einheitliche Terminologie schaffen, bevor mit dem Unterrichten der Intonationsphrasierung begonnen werden kann, da dies die Basis für alle weiteren prosodischen Merkmale darstellt. Intonationseinheiten sollten zusammen mit Flüssigkeitsphänomenen unterrichtet werden, die der L1 der Lerner angepasst werden sollten. Ein Fokus sollte insbesondere auf die Pragmatik-Prosodie-Schnittstelle gelegt werden, indem die Prosodie immer in einem Kontext untersucht und unterrichtet wird. Dabei kann man sich Sprachanalysesoftware bedienen und Aufnahmen von Studenten und Modelsprechern durchführen lassen, damit die Lerner nuancierte, kontextabhängige Bedeutungsunterschiede bei Muttersprachlern erkennen und lernen, diese nachzuahmen. Sowohl die unterschiedlichen Varietäten des Englischen als auch Sprachstilunterschiede sollten im Englischunterricht Berücksichtigung finden.

Im Folgenden werden – unter Einbeziehung der Forschungsfragen – die wichtigsten Erkenntnisse dieser Studie noch einmal zusammengefasst.

Kapitel 6

Zusammenfassung und Ausblick

Das letzte Kapitel fasst die wichtigsten Erkenntnisse des Dissertationsprojekts für die Lernerprosodieforschung, die Fremdsprachenforschung sowie die Zweitsprachenerwerbsforschung zusammen. Insgesamt hat die vorliegende Studie dazu beigetragen, unser Verständnis von Prosodie und deren pragmatische Bedeutung in der muttersprachlichen und nicht-muttersprachlichen Sprache zu verbessern. Es wurde ein multifaktorielles Modell vorgeschlagen, das basierend auf bestehenden Theorien und durch empirische Unterstützung der vorliegenden Arbeit entwickelt wurde. Aus der Zweitsprachenerwerbsperspektive fielen die Ergebnisse der vorliegenden Studie größtenteils unter die Vorhersagen mehrerer aktueller Zweitsprachenerwerbstheorien, d.h. SLM (*Speech Learning Model*; Flege 1995), BV (*Basic Variety*; Klein und Perdue 1997) und LILt (Mennen 2015). Im Gegensatz zu den Vorhersagen der BV-Theorie verwendeten die Lerner die Prosodie, um eine Vielzahl pragmatischer Funktionen zu erfüllen, den Übergang zu berichteter/zitierter Sprache zu markieren und sich den Interviewern in ihrer Grundfrequenz anzunähern und anzupassen (*prosodic entrainment*). Um zu den ursprünglichen vier Forschungsfragen zurückzukehren, so scheinen die Ergebnisse der vorliegenden Studie darauf hinzudeuten, dass für die Mehrheit der Lerner mit allen drei L1-Hintergründen ein nicht-muttersprachliches Intonationssystem zu Grunde zu liegen scheint, was frühere Forschungen bestätigt (z. B. Ramírez Verdugo 2006b). Die Anwendung von Mennens Theorie (2015) auf Lerner aus drei verschiedenen L1-Gruppen ergab, dass sich die Lerner hinsichtlich der verteilenden, realisierenden und funktionalen Dimension der Intonation von der Muttersprache unterscheiden. Die Ergebnisse dieser Studie, die frühere Forschungen bestätigen sind folgende:

- Für L2-Lerner mit unterschiedlichem L1-Hintergrund gibt es universelle Muster beim Erwerb der L2-Prosodie, da Lerner ähnliche Abweichungen von der muttersprachlichen Norm zeigen.
- Der Grundfrequenzumfang ist in der Lernersprache für HL-Melodien enger.
- Das *F0 Level* für HL-Melodien ist in muttersprachlicher und nicht-muttersprachlicher Sprache ähnlich und weicht nicht signifikant von der Muttersprachlernorm ab.

- Es gibt Unterschiede im Grundfrequenzumfang je nach der Position der f0-Kontour in der Intonationseinheit.
- Hohe Grenztöne und LH-Melodien werden in der Lernersprache häufiger verwendet als in dem Output der englischen Muttersprachler.
- Lerner produzieren eine langsamere Sprechgeschwindigkeit.

Die aktuelle Studie war insofern innovativ, als dass sie die Sprache von Lernern mit drei verschiedenen L1-Hintergründen im Rahmen der Annahmen des LILt-Frameworks verglichen hat und eine prosodisch-pragmatische Analyse mit einer Reihe prosodischer Merkmale kombinierte. Dies sind die neuen Erkenntnisse der vorliegenden Studie:

- Es gibt für alle drei Lernergruppen (richtungsmäßig gleiche) Unterschiede in Bezug auf die Verteilung und die Realisierungsdimension, die hauptsächlich auf Zweitspracherwerbsprobleme und einige L1-Transfereffekte zurückzuführen sind. Der Grad der Abweichung von jeder Intonationsdimension hängt aber von der Erstsprache, dem Sprachniveau und dem Geschlecht der Lerner ab.
- Es wurden geschlechtsspezifische Unterschiede bei fast allen prosodischen Variablen, insbesondere bei der Verteilung und Häufigkeit der Intonationseinheiten festgestellt.
- Es gibt signifikante melodiebasierte Unterschiede im Grundfrequenzumfang: Alle Lerner weisen Probleme bei der Realisierung und Verteilung von Melodien mit hohem Register (*high register tunes*) auf (abgesehen davon, dass für HL-Melodien eine engere Grundfrequenzspanne erzeugt wird). Die LH- und HH-Melodien aller Lerner sind extremer, d.h. höher im *f0-Level* und breiter in der *f0-span*.
- Form-Bedeutungsbeziehung (*form-meaning relationship*): qualitative Untersuchungen der sozio-pragmatische Funktionen von hohen Grenztönen und LH-Melodien haben gezeigt, dass Lerner diese verwenden, um Unsicherheit und Kohäsion zu signalisieren, jedoch auch als pragmatischen oder zwischenmenschlichen Marker (*interpersonal marker*). Daher können hohe Grenztöne als Indikator für das Sprachniveau der Lerner angesehen werden. Sie wurden dementsprechend auch als Beispiele für einen prosodischem Drift (*prosodic drift*) und eine prosodische Erkundungsphase (*prosodic exploration phase*) vorgeschlagen. Gleichzeitig können jedoch auch mögliche Vermeidungsstrategien als Erklärung dienen.

Auf methodischer Ebene hat die vorliegende Studie die Vor- und Nachteile eines multifaktoriellen, korpusbasierten Ansatzes zur Untersuchung der Effekte mehrerer extralinguistischer Faktoren auf die Lernerprosodie aufgezeigt. Trotz der Nachteile der für die vorliegende Studie ausgewählten Korpora boten diese wertvolle Einblicke in die prosodische Produktion und den Zweitsprachenerwerb, die in der zukünftigen Forschung mit anderen Korpora oder detaillierteren Analysen mit den hier verwendeten Daten reproduziert werden können. Die vorliegende Studie hat auch gezeigt, dass es möglich ist, quantitative und qualitative Analysen mit einem solchen Korpus zu kombinieren. Während die quantitative Analyse gezeigt hat, dass das Uptalk-Merkmal auf einen nicht-muttersprachlichen Output hinweist, was für alle drei Lernervarietäten gleichermaßen bewiesen wurde und die meisten Intonationsforscher zu dem Schluss geführt hat, dass dies mit einem höheren Grad an Unsicherheit seitens der Lerner korreliert, konnte die qualitative Analyse zeigen, dass Uptalk auch für eine Vielzahl von pragmatischen Funktionen verwendet werden kann, die den Kommunikationserfolg verbessern können. Dies zeigt, dass eine quantitative Analyse allein kein vollständiges Bild der Verwendung von hohen Phrasen- und Grenztönen von Nichtmuttersprachlern liefert. Weitere methodische Beiträge der Studie sind nachstehend aufgeführt:

- Es lohnt sich, hohe Töne mit detaillierteren pragmatischen Funktionen genauer zu untersuchen (siehe auch Staples und Fernández 2021), da nicht alle hohen Töne in der Lernersprache gleich sind. Die Entwicklung von IPrA könnte detailliertere Analysen von Tönen ermöglichen.
- Es kommt darauf an, mit welcher muttersprachlichen Varietät Lerner verglichen werden. Sprecher des amerikanischen Englisch aus dem Süden der USA weichen (in dieser Studie) am stärksten von der Prosodie der britischen Muttersprachlernorm ab und weisen in der Verteilung von intonatorischen Phrasierungsparametern und Melodien mit hohem Register größere Ähnlichkeiten mit der Sprache der Lerner auf.
- Die Kombination von intonatorischen Phrasierungsmaßnahmen mit Flüssigkeitsmaßnahmen (*fleuncy measures*) ist ein lohnendes Unterfangen (kein neuer Ansatz).
- Die vorliegende Studie weist auf die Schwierigkeit hin, die prosodische Verbesserung mit nur einem Kompetenzindikator zu messen (mehrere Indikatoren könnten aufschlussreicher sein).
- Die Kombination mehrerer prosodischer Merkmale liefert ein vollständigeres Bild der Abweichungen entlang der Intonationsdimensionen.

Wenn man die drei Lerngruppen miteinander vergleicht, kann nun abschließend gesagt werden, dass sie alle eine systematische Entwicklung in Richtung ihres Ziels zeigen. Insgesamt zeigen die Ergebnisse, dass der L1-Hintergrund eines Lerners vernachlässigbar erscheint und dass der Produktion von hohen Tönen und Melodien (Realisierungs- und Verteilungsdimension) und deren pragmatischen Effekten mehr Aufmerksamkeit im englischen Fremdsprachenunterricht geschenkt werden sollte. Hierbei scheint es sich um einen Entwicklungsfaktor zu handeln, den alle Lerner gemeinsam haben. Die Ergebnisse stützen die Annahme, dass sich hohe Töne langsamer entwickeln und schwieriger zu produzieren sind als tiefe (Ohala und Ewan 1973; Crystal 1986; Grosser 1997; Snow und Balog 2002; Mennen et al. 2010). Die vorliegende Studie hat auch gezeigt, dass es sich bei dem Erwerb von Prosodie um einen mehrstufigen Prozess handelt, der vom Erwerb (*attainment*) anderer prosodischer Merkmale abhängt, d.h. die Entwicklung der Intonationseinheiten beeinflusst die Produktion von Tönen und Melodien und muss daher (für bestimmte L1-Gruppen) zuerst im Sprachunterricht behandelt werden.

Anregungen und Ideen zur weiteren Forschung beinhalten unter anderem den Hinweis, dass Prosodie mit so vielen linguistischen Bereichen (z.B. Segmente, Flüssigkeit, Pragmatik, etc.) wie möglich kombiniert und vermehrt extralinguistische Faktoren in zukünftigen Studien mit einbezogen werden sollten. Nonverbale Signale (Gestik, Mimik, Handzeichen, etc.) konnten in der Analyse des vorliegenden Projekts auf Grund des fehlenden Videomaterials leider nicht mit berücksichtigt werden. Diese könnten in zukünftigen Studien jedoch Aufschluss über weitere kulturelle Unterschiede zwischen den Lernergruppen geben. Aus theoretischer Sicht bekräftigt die vorliegende Studie die Notwendigkeit einer Weiterentwicklung von L2-Theorien und -Modellen für die prosodische Akquisition. Diese sollten präzise Vorhersagen auf der Grundlage der Vielschichtigkeit des L2-Prosodie-Erwerbs enthalten und die Sprachniveaus der Lerner sowie das muttersprachliche Zielmodell berücksichtigen, an welches sie sich annähern möchten.

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Karin Puga

Gießen, 08.12.2020 ……………………………..

(Name & Unterschrift)