

# Parentheticals in spoken Indian and Sri Lankan English

Julia Degenhardt 

Department of English, Justus Liebig University Giessen, Giessen, Germany

## Correspondence

Julia Degenhardt, Department of English, Justus Liebig University Giessen, Otto-Behaghel-Straße 10B, 35394 Giessen, Germany.

Email: [Julia.Degenhardt@anglistik.uni-giessen.de](mailto:Julia.Degenhardt@anglistik.uni-giessen.de)

There are no funders to report for this submission.

## Abstract

While an increasing number of studies into the pragmatics of world Englishes indicate that sociobiographic factors—such as the speakers' age or gender—influence pragmatic choices, most empirical investigations do not include sociobiographic information beyond said speaker characteristics. This study investigates parenthetical *I assume/believe/feel/guess/suppose/think* in the spoken parts of the Indian and Sri Lankan English components of the International Corpus of English to answer the question of to what extent parenthetical function, that is, expressing either the speaker's opinion or insufficient knowledge, is influenced by structural, contextual and sociobiographic factors. Based on 1265 parentheticals, the results of multifactorial statistical analyses indicate that the speakers' educational background and additional languages spoken at home are important predictors for the choice of parenthetical function. Therefore, the study calls for the inclusion of wide-ranging sociobiographic factors (and combinations thereof) in the description of pragmatic speaker choices in world Englishes.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *World Englishes* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Research interest in the pragmatics of (spoken) South Asian Englishes has recently seen an increase. Often comparing Indian English (IndE) and Sri Lankan English (SLE) to their historical input variety British English (BrE), corpus-based research in said linguistic field includes conversation analysis (Kraaz & Bernaisch, 2022, on backchannels; Revis & Bernaisch, 2020, on pauses), speech acts (Degenhardt, 2021, on requests; Degenhardt & Bernaisch, 2022, on apologies; Funke, 2022, on thanking) and indications of speaker commitment towards the utterance (Funke & Bernaisch, 2022, on intensifying and downtoning) amongst others. While some of these studies can attest the importance of sociobiographic factors for the speakers' choice between pragmatic alternatives, for the sake of comparability between varieties, additional speaker information available only in some of the metadata of individual varieties in the International Corpus of English (ICE), such as the speakers' educational background or their occupation, is often neglected.

Therefore, this study takes on a two-fold approach to the description of pragmatic speaker choices of parenthetical comment clause function in South Asian Englishes by (a) including information on the speakers' home languages and additional languages for both IndE and SLE and (b) investigating the importance of the speakers' education and occupation to predict different comment clause functions in SLE only. Taking into account both private and public dialogues of the spoken part of ICE-India (ICE-IND) and ICE-Sri Lanka (ICE-SL), this study offers a multifactorial statistical perspective on comment clauses annotated for various structural, contextual and sociobiographic predictor variables by employing a generalised linear mixed model tree (glmertree; Fokkema et al., 2018, 2021) and a Random Forest (RF; Breiman, 2001) analysis.

This paper first introduces a selected number of previous research into parentheticals/comment clauses as well as the pragmatics of IndE and SLE. Section 3 presents the data and annotation used in this study in more detail and highlights methodological implications. Section 4 offers quantitative and qualitative analyses of the results. Section 5 discusses the results against the background of the research questions and, finally, Section 6 concludes the present study while providing an outlook for future research.

## 2 | THEORETICAL BACKGROUND

### 2.1 | Parentheticals

The object of investigation of this study is parenthetical expressions, more specifically certain realisations of one type of parentheticals, namely, comment clauses. Parentheticals are often defined as a 'linguistic entity which is linearly integrated in another linguistic structure but is unrelated to the surrounding linguistic material ...' (Dehé, 2014, p. 1). Apart from that definition, researchers' understandings of potential members of the category of parentheticals vary greatly from including single word units to full clauses, from various clausal realisations (such as main clauses, adverbial clauses, relative clauses and so on) to question tags, interjections or filled pauses. Previous research has shown that parentheticals are attached to host constructions, but they can be taken out of the host without changing their meaning (Biber et al., 1999, p. 1067; de Vries, 2012, p. 153; Dehé, 2014, p. 1). While this may hold true from a truth-conditional semantics perspective, from the point of view of pragmatics that the present paper takes on, it is highly doubtful that all types of parentheticals can be excluded from the overall construction without a change in meaning. Especially with regard to comment clauses—which derive their name from their function to comment on the rest of the sentence or utterance—there is a clearly perceivable difference between, for instance, *He moved to Spain* and *I think he moved to Spain*. The former is a straightforward statement, whereas the latter includes a comment clause (underlined) to signal a more tentative point of view. Thus, some types of parentheticals do indeed express 'speaker-oriented content' (Kluck, 2011, p. 229), reflect speaker attitude (Brinton, 2008, p. 18) or serve as mitigating devices (Fraser, 1980; Kaltenböck, 2010; Mittwoch, 1979).

Comment clauses in particular have frequently been investigated in Inner Circle varieties (Kachru, 1992), and comment clauses, too, come in various realisations. The most frequently occurring lexical comment clause type is *I think* (Kaltenböck, 2010, p. 250), which fulfils various pragmatic functions in previous studies that differ slightly in terms of denomination and range. Much of that early research focusses on qualitatively investigating academic writing or is rather intuition-based. Following Arndt (1987, p. 27), *I think* can express either personal attitudes/opinions, as in '*I think he's a fool*' (Kaltenböck, 2010, p. 239), or insufficient evidence/belief, as in '*I think he's a carpenter*' (Kaltenböck, 2010, p. 239). Persson (1993) first makes a four-way distinction between cogitation, probability, pure opinion and subjective evaluation. Cogitation represents the literal meaning of *I think* (as in *I will think about it*), probability has 'to do with the verifiability or falsifiability of the proposition' (Persson, 1993, p. 6), in which case *I think* can be similar to *I believe*. Pure opinion *I think* 'accompan[ies] propositions which do not involve the assessment or judging of probabilities' and often occurs when speakers express 'some kind of obligation or necessity' (Persson, 1993, p. 7). The fourth type of *I think*, subjective evaluation, is used to express the speaker's subjective impression in a given situation. However, Persson (1993, p. 9) himself states that it might be difficult to distinguish between probability and pure opinion *I think* and thus adopts a three-fold distinction that combines the two types under the category opinion, with cogitation and impression being the other two categories.

More recent studies into comment clause realisations with parenthetical *I think* are corpus-based, yet they differ greatly in the number of instances examined. Mindt (2003), for example, examines 1724 instances of *I think* in the British component of ICE (ICE-GB) and concludes that *I think* can take on the function of a discourse marker. Kaltenböck (2010) tries to find a way that includes both quantitative and qualitative evaluations of 379 occurrences of *I think* in ICE-GB. He identifies four pragmatic functions of *I think*. First, the shield function *I think* expresses 'the speaker's tentativeness, lack of certainty or commitment with regard to the truth value of its host' (Kaltenböck, 2010, p. 242). Second, the approximator function is quite similar to the shield function, but *I think*, as an approximator, 'indicates lack of precision of the term within its scope' (Kaltenböck, 2010, p. 248) not of the entire proposition. Therefore, Kaltenböck also highlights the importance of the co-text of *I think* in order to distinguish between the first two functions, as well as comment clause scope and its place in the utterance by stating that 'there are certain positions that are exclusively linked to phrasal scope' (Kaltenböck, 2010, p. 245), namely, if *I think* occurs between a preposition and its respective NP or between an NP and its determiner or modifier. Third, the structural function refers to instances of *I think* that are either phonetically reduced or integrated into the host construction via prosody (Kaltenböck, 2010, p. 250). Last but not least, the booster function stands in contrast to the shield function in that it enhances the speaker's commitment to the host construction's truth value. Both structural and booster functions are evaluated by means of, for instance, pitch differences as shown in spectrograms.

This study tries to find a common ground of the studies introduced above and thus makes a distinction between (a) the literal meaning of *I think* (Persson's cogitation category), (b) *I think* as a tentative expression of opinion (similar to Persson's combined opinion category) and (c) *I think* as an expression of insufficient knowledge (similar to Arndt's insufficient evidence/belief category). Kaltenböck's distinction between shields and approximators can be reflected in various levels of independent variables such as SCOPE or PLACE (see Section 3 for more information). Furthermore, while other comment clause forms, such as *I believe* or *I assume*, have not been extensively researched due to their relatively infrequent use compared to *I think*, speakers might opt to use them in similar, but not identical, co-text and contexts as *I think*. Hence, this study not only focusses on *I think* but also includes other comment clause forms to investigate if there are patterns that drive the usage of another form when *I think* could have been used as well. It will be particularly interesting to see if there are variety-specific preferences for a comment clause form other than *I think* to speculate about possible first-language influences on pragmatic speaker choices.

## 2.2 | The pragmatics of Indian and Sri Lankan English

IndE and SLE are varieties of English that arose in the context of the British colonisation of the Indian subcontinent. Other South Asian varieties of English include Bangladeshi English, Maldivian English, Nepali English and Pakistani

English. IndE—as the oldest and furthest developed South Asian variety following Schneider's (2007) dynamic model—is often assumed to exert epicentral influence on neighbouring varieties (Gries & Bernaisch, 2016; Mukherjee, 2008). In Sri Lanka, the English language underwent 'institutionalisation processes via missionary schools and language policies under British administration' (Bernaisch, 2015, p. 1) that are comparable to that of English in India yet detached from those processes in India. English is not equally well represented throughout both countries. Urban centres of India and Sri Lanka are often also centres of higher education in which English is one of the main languages. Therefore, a small minority of people also learn English as a first language, amongst other indigenous first languages. Hence, factors such as the speakers' educational background, languages spoken at home or additional languages learnt are assumed to have an influence on speakers' usage of English.

Research into IndE and SLE has predominantly focussed on structural features and attested structural nativisation, sometimes even endonormative stabilisation of both varieties (e.g. Lange, 2012, on IndE syntax; Bernaisch, 2015, on the lexis and lexicogrammar of SLE; Gries & Bernaisch, 2016, on the dative alternation in South Asian Englishes). Regarding the pragmatics of spoken IndE and SLE, many of the recent studies include a variety of sociobiographic factors, the influence of which on the pragmatic response variable is often tested multifactorially.

Investigating the choice between filled and unfilled pauses (e.g. *uhm* and silence) in ICE-GB, ICE-IND and ICE-SL, Revis and Bernaisch (2020, p. 142) extract 88,212 examples and annotate them according to sociobiographic speaker data (VARIETY, AGE and GENDER) and a couple of structural factors such as the WORDCLASS (either lexical or functional) of the word following the pause or the type of DISCOURSE the pause occurred in (dialogue vs monologue). With the help of a conditional inference tree (CTree) and a generalised linear mixed-effects model (GLMM), the authors conclude that AGE and GENDER are important predictors of the choice between the use of filled and unfilled pauses, yet both factors are no longer significant once a possible speaker-bias is statistically controlled for in the GLMM.

Kraaz and Bernaisch (2022) model roughly 3200 instances of backchannels found in the spoken parts of ICE-GB, ICE-IND and ICE-SL multifactorially (via CTree and RF analyses) according to AGE and GENDER of both the backchannel user as well as of the speakers directly contributing to the backchannel situation and the type-token ratio (TTR) of both backchannel user and said other speaker(s), amongst other variables. They conclude that both IndE and SLE speakers use backchannels more frequently than BrE speakers, but the most important variable influencing backchannel frequencies is TTR, only followed by GENDER and AGE, thus advocating for the inclusion of not only sociobiographic but also structural factors. The study furthermore attests to 'a backchannel that replicates the last nominal head used by the preceding interlocutor' (Kraaz & Bernaisch, 2022, p. 239), a strategy unique to the investigated two South Asian varieties.

Similar conclusions are drawn by Funke and Bernaisch (2022), who multifactorially investigate differences in frequencies and forms of intensifiers and downtoners across the same three varieties, with differences between said varieties boiling down rather to structural predictors (such as TTR) than sociobiographic ones (Funke & Bernaisch, 2022, p. 59ff.). Speakers of all three varieties use expressions that seem to represent variety-preferential strategies of intensifying and downtoning, presumably based on strategies taken from indigenous first languages in the case of IndE and SLE (Funke & Bernaisch, 2022, p. 33). Hence, the inclusion of additional predictors such as additional languages spoken by the respective speakers might be worthwhile.

Also focussing on ICE-GB, ICE-IND and ICE-SL, Funke (2022) studies the influence of various structural and sociobiographic factors on thanking strategies in the three varieties. Structural factors in the choice between *thank* and *thanks* include the thanking strategy's POSITION within the speaker's turn as well as the presence of an INTENSIFIER, a BENEFACOR or REASON. Sociobiographic factors include VARIETY, speaker's AGE and GENDER, and the context's FORMALITY. Using a CTree, Funke concludes that VARIETY, AGE as well as the presence of a BENEFACOR and an INTENSIFIER predict the choice of *thank* and *thanks*. For instance, 'IndE and SLE speakers will choose *thanks* relatively more frequently if no or an unspecified INTENSIFIER is part of the thanking formula and the BENEFACOR takes the form of an *endearment*, a *group*, or an *institution*' (Funke, 2022, p. 145, italics and capitals in original). In addition to that, Funke's (2022) RF anal-

ysis indicates that an interaction of VARIETY and the other variables in her study significantly influence the choice of *thank* over *thanks*. Regarding the interaction between VARIETY and AGE, Funke's (2022, p. 146) analysis suggests that older BrE and IndE speakers often choose *thank* over *thanks*, whereas SLE speakers are more likely to choose *thank* over *thanks* in the younger age group, hence underlining the explanatory value of interaction predictors.

Degenhardt and Bernaisch (2022) look at apologies in IndE and SLE compared to BrE to find possible variety-specific apology patterns. Concentrating on the choice between overwhelmingly more frequent realisation patterns involving *sorry* versus other forms of apologies (such as *apologise* or *forgive*), they study the influence of VARIETY, GENDER, GROUPGENDER, AGE, INTENSIFIER, SETTING, TYPE, TOPIC and TTR with a CTree and an RF analysis. They conclude that apology realisations seem to depend on the speakers' AGE and GENDER in that, for instance, women appear to be apologising more frequently than men and with various apology strategies—which are, however, not unique in either of the varieties (Degenhardt & Bernaisch, 2022, p. 219).

Comment clauses have not yet been investigated empirically in South Asian varieties of English, yet at least one other type of parentheticals has been studied in IndE: Lange and Leuckert (2020) study tag questions in the private dialogues data of ICE-IND, with special recognition of gender differences in the speakers' usage. Their study concludes that female speakers seem to use question tags much more frequently than male speakers do, and their findings suggest a certain degree of pragmatic nativisation of IndE tag questions (Lange & Leuckert, 2020, p. 89).

Against the background of earlier studies into IndE and SLE, it can be assumed that South Asian varieties exert variety-specific or pan-South Asian realisation patterns on structural linguistic levels at least to some extent. Moreover, the various studies into the pragmatics of South Asian English varieties introduced above show that whether or not sociobiographic speaker factors, such as speaker AGE or GENDER, play an important part in determining variety-specific or variety-preferential pragmatic speaker choices varies with the object of investigation. Especially following other studies into parentheticals in South Asian Englishes (Lange, 2012; Lange & Leuckert, 2020) that attested to the importance of said sociobiographic factors (as well as structural factors), it seems worthwhile to consider these factors when investigating comment clauses.

### 2.3 | Research gaps and questions

The theoretical background laid out above indicates a couple of research gaps. First, previous studies into parentheticals in general and comment clauses in specific have mostly concentrated on Inner Circle varieties, so there is still a lot of research to be conducted in other varieties of English. Second, and more specifically regarding South Asian Englishes, while research into the pragmatics of South Asian Englishes may be on the increase, there is still a variety of pragmatic objects of investigation as well as a variety of additional speaker variables—such as the speakers' educational background, occupation or additional languages—left worthy of attention. Third, research into (the pragmatics of) world Englishes in general is lacking a great deal of empirical studies that involve multifactorial (instead of monofactorial) analyses, the former of which are essential to a more all-encompassing understanding of the effect of the interplay of various independent variables on the response variable of interest. Therefore, based on the theoretical background laid out above, the present study aims to answer the following research questions:

1. Are there quantitative differences regarding comment clause usage between the two South Asian varieties, either in terms of overall frequencies or with regard to preferences towards any of the comment clause functions?
2. In how far is comment clause function influenced by structural, contextual and sociobiographic factors in IndE and SLE, and which of these factors, or a possible interplay of these factors, determines comment clause speaker choices best in each of the varieties?
3. Does the inclusion of additional sociobiographic factors contribute significantly to a better understanding of pragmatic speaker choices?

## 3 | METHODOLOGY

The study at hand uses the private and public dialogue data of the spoken part of the IndE (Shastri, 2002) and SLE (Bernaisch et al., 2019) components of ICE. Despite presumably similar compilation processes of individual ICE components, the spoken part of ICE-SL includes a considerable amount of additional sociobiographic speaker metadata, which could not be taken into account in any of the studies introduced above due to their focus on comparing SLE to IndE and BrE. Therefore, this study concentrates on (a) a comparison of IndE and SLE to show possible overall South Asian patterns and (b) SLE only to investigate the influence of additional sociobiographic factors that are not recorded in the IndE component.

The latter is a valuable addition to the former to state whether the implementation of additional sociobiographic factors brings about a deeper understanding of pragmatic choices in postcolonial varieties. With pragmatics being the study of contextual speaker meaning (Yule, 1996), pragmatic investigations—by definition—take into consideration the language user and can, thus, only profit from as much sociobiographic and contextual information as possible.

Prior to data extraction, foreign speaker material such as utterances by BrE speakers were automatically excluded from the corpus data to guarantee the inclusion of IndE and SLE only. With the help of the corpus linguistic software package AntConc (Anthony, 2018), the data were then searched for the following expressions: *I assume, I believe, I feel, I guess, I suppose, I think*. A total of 1287 parentheticals were extracted from the corpus and manually cross-checked for their function as a comment clause. Hence, instances of, for example, *I feel smart* (ICE-SL:S1A-050#68:1:A) were excluded from the data. Literal uses of *I think* or *I believe* (that is the first category laid out above which represents literal uses of parenthetical expressions) were also excluded from the data set as they do not represent purely pragmatic functions. Consequently, a total of 1265 parenthetical comment clauses in both South Asian varieties were annotated for the sociobiographic, contextual and structural independent variables outlined further below to predict either of the two levels of the dependent variable FUNCTION, namely, *expressing opinion* or *expressing insufficient knowledge*. Many instances could be categorised according to the general rule that once the truth or correctness of an utterance involving a comment clause realisation could be attested from an objective point of view (example (1)), it would be categorised as expressing insufficient knowledge, whereas comment clauses expressing personal attitudes or assumptions that cannot easily be verified were categorised as expressing an opinion (example (2)).

- (1) **I think** now Sri Lankan Airlines is also flying from that <, > very <, > stupid airport. (ICE-SL:S1A-027#415:1:B)
- (2) ... **I think** the flavour is not that nice. (ICE-SL:S1A-046#303:1:A)

One might state that a binary distinction is not enough for all possible pragmatic functions comment clauses such as *I think* can fulfil. However, the decision to make FUNCTION a binary variable and not include a more fine-grained distinction between various functions of comment clauses is based on a couple of considerations. First, previous studies (e.g. Arndt, 1987) were based on similar binary distinctions or choosing a binary distinction that best represents the common ground of previous studies, thus ensuring that this study is more or less comparable to previous studies. Second, though more recent studies, such as Kaltenböck's (2010) study, include at least four pragmatic functions when investigating spoken data, two of them (structural and booster function) are categorised on the grounds of spectrogram evaluations of intonation patterns. While this study looks at spoken language as well, involving such analyses of intonation patterns is outside the scope of a corpus-based study. Third, the focus of this study lies in describing overall patterns in IndE and SLE speaker choices as influenced by a wide range of factors, including many (additional) sociobiographic factors, to account for possible sociolinguistic variation in South Asian Englishes. Hence, from a statistical perspective, a more fine-grained distinction of comment clause FUNCTION and thus the annotation of relatively infrequent patterns in FUNCTION choice seems counterproductive in the light of the number of independent variables taken into consideration.

There are three different types of independent variables in the present study, contextual, structural and sociobiographic, with a predominant focus on the latter two. Contextual predictors include information on possibly influential factors surrounding the communicative situation that do not have to do with the speakers themselves. The present study includes the contextual predictor *SETTING*, which deals with the communicative context as either *private* versus *public* following the ICE corpus design.

Structural predictors are concerned with the immediate comment clause co-text:

- **FORM**: the respective comment clause realisation, namely, comment clauses involving *assume* vs *believe* vs *feel* vs *guess* vs *suppose* vs *think*,
- **PLACE** of comment clause within the utterance: utterance *initial* vs *medial* vs *final*,
- **SCOPE** of comment clause: *phrasal* vs *clausal*,
- **REPETITION**: whether the same comment clause expression is repeated by the speaker in the immediate co-text, yes (**I guess I guess** she can because you know Kim Kardashian is doing it as it is (ICE-SL:S1A-066#208:1:A)) vs no, though not as a form of self-repair,
- **COMPOUND**: whether the comment clause occurs in combination with other parentheticals or hedges, yes (... he's the I think uh **you know according to my knowledge** one of the best presidents (ICE-SL:S1A-053#78:1:B)) vs no,
- **DYSFLUENCY**: whether the comment clause occurs in combination with at least one dysfluency, such as filled and unfilled pauses or false starts (Götz, 2013), yes ([...] I think <„> **uh** <„> think what she's hinting at is to ask about me playing the piano (ICE-SL:S1A-074#203:1:B)) vs no.

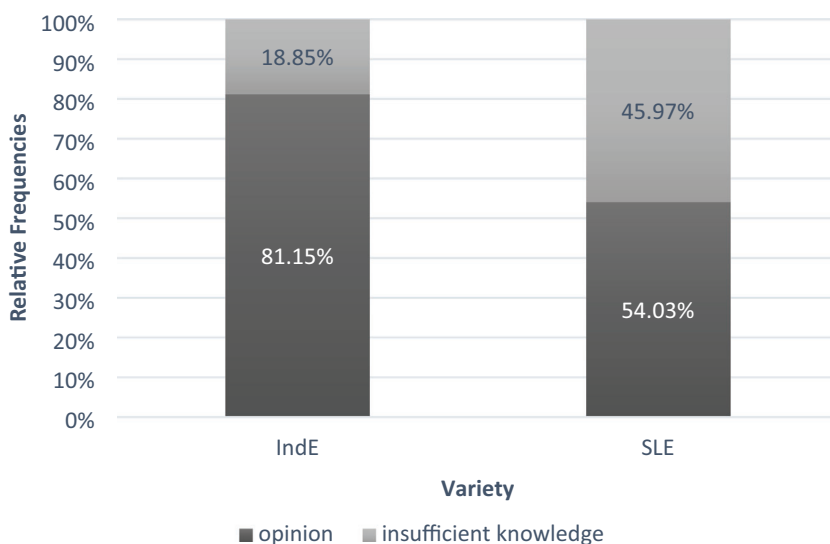
Last but not least, the sociobiographic predictors included in this study are:

- **VARIETY**: *IndE* vs *SLE*,
- **AGE** of speaker: *younger* ( $\leq 26$ ) vs *older* ( $> 26$ ),
- **GENDER** of speaker: *female* vs *male*,
- **GROUP\_GENDER**: whether the interlocutors are of the *same* or of a *different* gender than the speaker who uses a comment clause,
- **HOME\_LANGUAGE(S)**: other languages spoken inside the speakers' homes (following available metadata), namely, *Dravidian* vs *Indo-European* vs *Dravidian\_Indo-European* vs *English* vs *English\_Indo-European*,
- **ADDITIONAL\_LANGUAGE(S)**: other languages spoken by the speakers, at home and elsewhere, that is *Dravidian* vs *Indo-European* vs *Dravidian\_Indo-European* vs *English* vs *English\_Indo-European* vs *English\_Dravidian* vs *English\_Dravidian\_Indo-European*.

For the investigation of only SLE, 696 comment clauses after disambiguation were annotated for the dependent and independent variables above, and two additional predictors, the information of which was retrieved from the ICE-SL metadata:

- **EDUCATION** of speaker: *GCE\_Advanced* vs *Higher National Diploma (HND)* vs *BA* vs *MA* vs *PhD*,
- **OCCUPATION** of speaker: *student* vs *professional*.

Some of these variables need further explanation. The distinction between phrasal and clausal **SCOPE** is based on Kaltenböck (2010, p. 238), who states 'that the scope of *I think* is not always clausal, ... covering the entire host clause, but may also be phrasal, [for example] singling out parts of the host construction'. Hence, a corpus example of clausal scope would be ... *I think you didn't try it* (ICE-IND:S1A-029#90:1:A), an example of phrasal scope would be *Yes I suppose so* (ICE-IND:S1A-055#165:1:A). Moreover, South Asian English speakers come with a wide variety of different home languages and additional languages, yet not all of them occur amongst speakers using a comment clause and too fine-grained distinctions between single languages are not feasible for statistical handling of the data. Therefore, regarding



**FIGURE 1** Relative frequencies of levels of FUNCTION. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

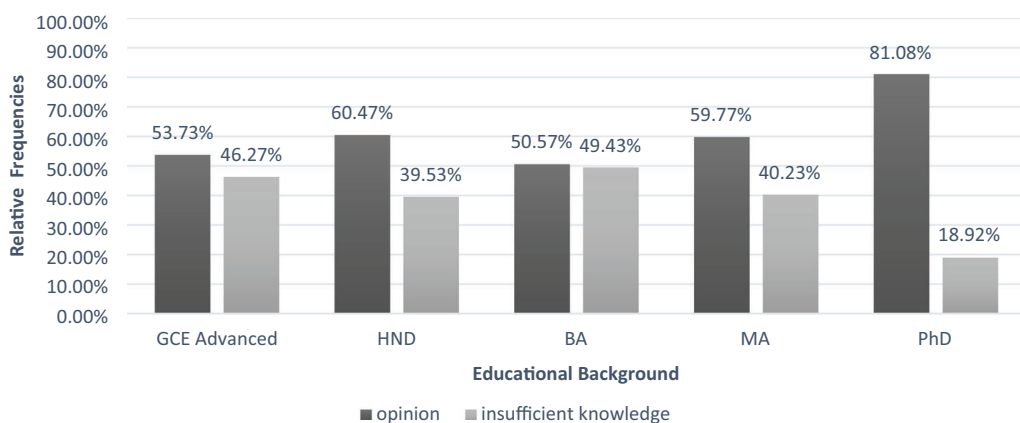
the predictors HOME\_LANGUAGE(S) and ADDITIONAL\_LANGUAGE(S), languages such as Tamil or Telugu were grouped under Dravidian languages, Hindi and Sindhi (amongst others) were grouped under Indo-European languages, and combinations of Dravidian and Indo-European languages with each other as well as with English were noted for those speakers who indicated to have more than one home language or more than one additional language. While generally speakers with a combination of, for instance, English + Dravidian + Indo-European home languages may be featured in the whole ICE-SL data set, they were not attested in the comment clause data set, which is why the annotation scheme differs slightly between HOME\_LANGUAGE(S) and ADDITIONAL\_LANGUAGE(S).

In a similar vein, concerning the predictor OCCUPATION, all possible professions were subcategorised into the binary distinction between student and professional. Though many of the younger speakers are students, some of them are also professionals, and while many of the older speakers are professionals, some of them are also still students, so annotating for both AGE and OCCUPATION is useful to see if there are differences amongst the speakers that might be influenced by their occupational background.

## 4 | ANALYSIS

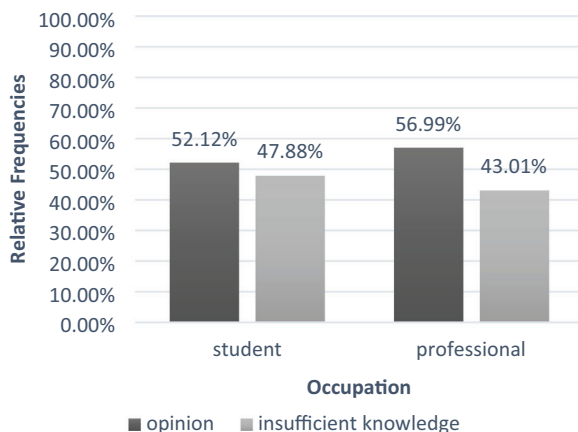
In order to answer the question of whether there are variety-specific preferences between IndE and SLE speakers, a glance at descriptive statistics provides a first tentative impression of the data. Figure 1 shows that with regard to the dependent variable, speakers of IndE preferably use the comment clause forms under investigation to express speaker opinion over insufficient knowledge, whereas SLE speakers use both functions in a more balanced manner. In general, however, 86.73% of all comment clauses were realised with *I think*, 9.78% with *I guess*, 2.65% with *I suppose* and *I assume/believe/feel* shared the remaining 0.84%, so the impression raised in Figure 1 is likely to be mostly due to the use of *I think*.

Normalised frequencies show that, in general, speakers with a GCE Advanced degree use more comment clauses than the other speaker groups and students use comment clauses slightly more frequently than speakers who are already in a professional context. Concerning the additional sociobiographic factors included only for SLE, the relative frequencies in Figures 2 and 3 indicate that EDUCATION and OCCUPATION could play a role in speakers' comment clause realisations. Speakers often slightly prefer to express opinion regardless of their educational background, with



**FIGURE 2** Relative frequencies of levels of FUNCTION according to EDUCATION in SLE. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

**FIGURE 3** Relative frequencies of levels of FUNCTION according to OCCUPATION in SLE. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

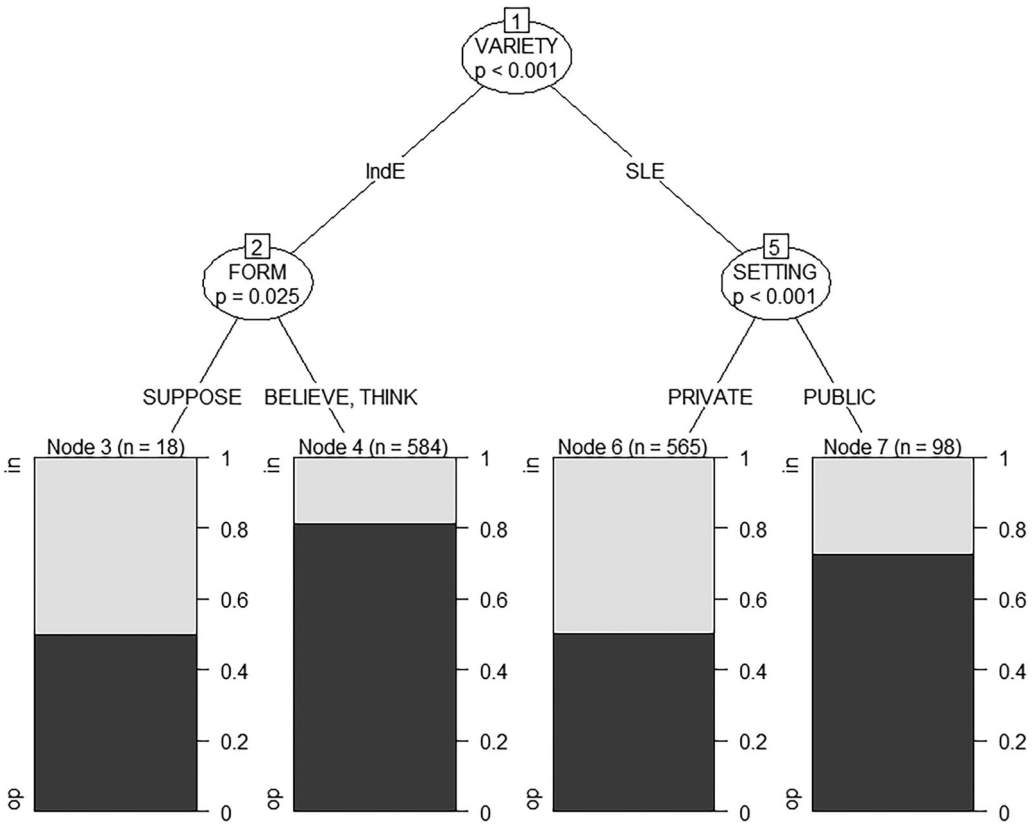


PhD holders even expressing opinion in about 80% of the cases (Figure 2). Contrary to that, regarding OCCUPATION (Figure 3), students do not seem to make fundamentally different choices than professionals.

However, this is only a monofactorial representation of the data with little to no explanatory power as towards which role each of the study's predictors plays when all the other predictors are being controlled for statistically. As Gries (2018, p. 295) points out

- (i) no phenomenon is monofactorial and (ii) even if one had a new monofactorial hypothesis of a phenomenon, it would still require multifactorial testing to determine either (a) whether it either adds anything to what we already know about the phenomenon (by statistically controlling for what we already know) or (b) whether it replaces (parts of) what we already know about the phenomenon.

Therefore, in order to answer the research questions of the study at hand, a glmertree was employed (Fokkema et al., 2018, 2021). Glmertrees are a mixture of traditional tree-based recursive partitioning models and GLMMs and thus improve on traditional tree-based models by combining the flexibility of tree-based models with the strengths of GLMMs (Bernaisch, 2022, p. 177). Hence, a glmertree not only includes tree-based recursive partitioning to model important (interaction) predictors automatically but can also account for the fact that many speakers are likely to



**FIGURE 4** Glmertree of most important predictors for choice of FUNCTION in IndE and SLE.

contribute more than just one data point to the data set (Bernaisch, 2022, p. 177). Furthermore, GLMMs make it possible to include fixed and random effects. Taking the present study as an example, fixed effects can be found with the independent variables REPETITION and PLACE of parenthetical in the utterance. Every possible scenario is covered by the various levels of both variables: repetitions do either occur or do not occur, and the parenthetical expression can be placed either in utterance-initial, medial or final position. A random effect of the present study—and probably of every study looking at authentic speaker data—is the speakers themselves as the data collected in ICE-IND and ICE-SL only represent a random sample of IndE and SLE speakers. Hence, the glmertree includes a random effect for speaker (META.ID) to also take into account possible idiosyncrasies of individual speakers (Heller, 2017, p. 85).

The model was fit in RStudio. The resulting model formula for the analysis of IndE and SLE looks as follows:  $\text{FUNCTION} \sim 1 \mid \text{META.ID} \mid \text{VARIETY} + \text{GENDER} + \text{GROUP\_GENDER} + \text{AGE} + \text{HOME\_LANGUAGE}(s) + \text{ADD\_LANGUAGE}(s) + \text{SETTING} + \text{FORM} + \text{PLACE} + \text{SCOPE} + \text{REPETITION} + \text{COMPOUND} + \text{DYSFLUENCY}$ . The model has a classification accuracy of 80.03%, which is significantly ( $p < 0.001$ ) better than the baseline model/no information rate at 66.32%, which would always predict the more frequently occurring level of the dependent variable (in this case, expressing opinion).

The model formula for the analysis of only SLE excluded the predictor VARIETY but included the two additional sociobiographic factors mentioned above:  $\text{FUNCTION} \sim 1 \mid \text{META.ID} \mid \text{EDUCATION} + \text{OCCUPATION} + \text{GENDER} + \text{GROUP\_GENDER} + \text{AGE} + \text{HOME\_LANGUAGE}(s) + \text{ADD\_LANGUAGE}(s) + \text{SETTING} + \text{FORM} + \text{PLACE} + \text{SCOPE} + \text{REPETITION} + \text{COMPOUND} + \text{DYSFLUENCY}$ . The SLE model has a classification accuracy of 75.42%, which is significantly ( $p < 0.001$ ) better than the baseline model/no information rate at 51.02%.

Figure 4 shows the glmertree for the choice of FUNCTION amongst IndE and SLE speakers. The first split (node 1) happens according to VARIETY, which is the most important predictor for the choice of expressing opinion or insuffi-

cient knowledge. Hence, speakers of IndE and SLE are influenced by different additional factors when it comes to their choice of the dependent variable. For IndE (left split), the next important decision is made with regard to comment clause FORM (node 2), namely, either *I suppose* or *I believe* and *I think*. Whereas the former is used to express both opinion and insufficient knowledge in a balanced manner (node 3), *I believe* and *I think* are preferably used by IndE speakers to express opinion (node 4). SLE speakers (right split) appear to make a difference between utterances taking place in the private and the public sphere (node 5). During private conversations (node 6), SLE speakers express opinion as well as insufficient knowledge in roughly 50% of the times respectively. In public settings, however, they prefer using comment clauses to express opinion (node 7).

A qualitative look at the data shows that amongst IndE speakers, *I suppose* is frequently used when collectively involving other speakers with the use of *we* or *our* no matter the level of FUNCTION (example (3)), whereas *I believe* and *I think* are predominantly used in utterances that are only about the speakers (example (4)).

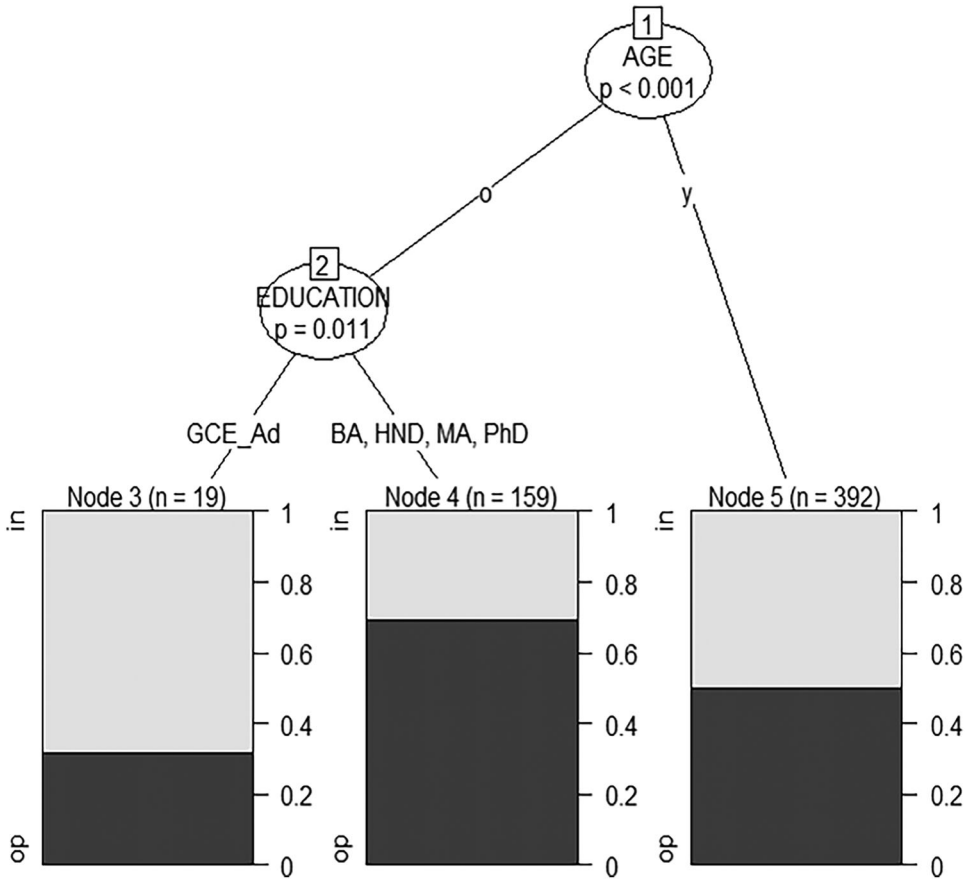
- (3) Could you tell us something about <,> Bengali <,> way of cooking (ICE-IND:S1A-007#14:1:A)  
Way of cooking <,> uh **I suppose** our staple food is fish and rice (ICE-IND:S1A-007#15:1:D)
- (4) Now gossip **I believe** is like humour <,> which is peculiar to every language (ICE-IND:S1A-080#6:1:B)

In SLE private conversations, speakers do not seem to prefer either of the two levels of FUNCTION, but qualitatively comment clauses expressing speaker opinion are often formulated rather directly, especially when asked for it by other interlocutors (example (5)), but comment clauses expressing insufficient knowledge are expressed more tentatively (example (6)). In public settings, for instance, courtroom or classroom discussions, SLE speakers prefer the use of expressing opinion.

- (5) What do you think <,> (ICE-SL:S1A-095#60:1:B)  
**I think** yeah <,> it's fine (ICE-SL:S1A-095#61:1:A)
- (6) **I think** she wanted to but I really don't know (ICE-SL:S1A-010#171:1:A)

Surprisingly, in the glmertree of both IndE and SLE, neither other contextual factors nor any of the sociobiographic factors apart from VARIETY seem to play a role in the choice of the dependent variable. However, SLE speakers make a clear distinction between private and public settings, and having access to sociobiographic metadata for SLE speakers that stand in direct connection with possible private versus public discourse (namely, OCCUPATION and EDUCATION) opens up the possibility for further investigation. Figure 5 shows that, in fact, the glmertree for SLE predicts AGE (node 1) and EDUCATION (node 2) to be determining factors for the choice of comment clause FUNCTION. Younger SLE speakers (node 5) do not show a preference for either level of FUNCTION, but older speakers with a GCE Advanced degree (node 3) show a surprising preference for expressing insufficient knowledge over expressing opinion, whereas it is the other way around for all speakers with a different educational background (node 4). A look at the data suggests that there are also slight qualitative differences in that speakers with HND or higher seem to express opinion with what could be an additional negative layer via the use of *I think*. Example (7) shows a discussion about the difference between books and their respective movies, and the part involving the comment clause could be interpreted to express that the speaker wants to opt out of the conversation or to indirectly ask the interlocutor to stop talking.

- (7) Maybe you've read the book <,> (ICE-SL:S1A-095#120:1:A)  
No (ICE-SL:S1A-095#121:1:A)  
No (ICE-SL:S1A-095#122:1:B)  
Yeah check it <,> (ICE-SL:S1A-095#123:1:A)  
**I think** it's high time I watched a movie now (ICE-SL:S1A-095#124:1:B)



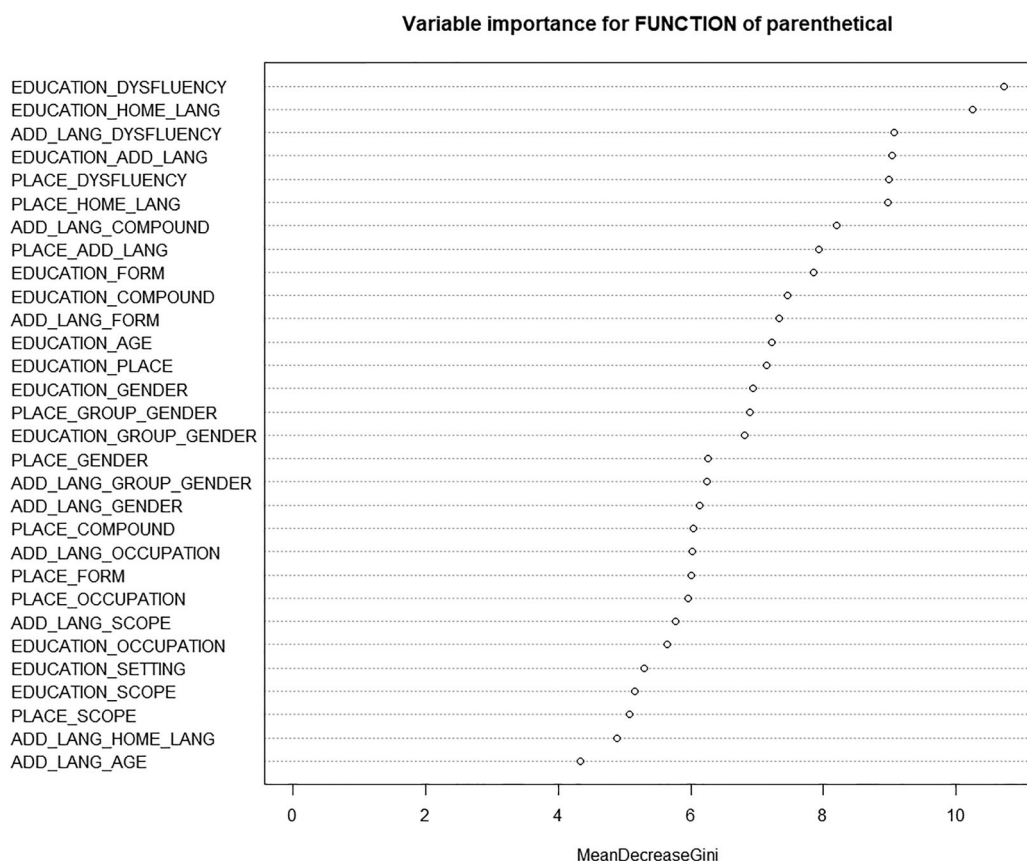
**FIGURE 5** Glmertree of most important predictors for the choice of FUNCTION in SLE.

However, a valid judgement of whether or not the utterance carries this underlying, rather negative flavour can only be made by further looking at intonation patterns and the like.

While the glmertree analyses of South Asian Englishes show that at least some of the predictors that were assumed to be important for the choice of expressing opinion or insufficient knowledge are indeed important, most of the predictors attested to be of importance in previous studies on the pragmatics of South Asian Englishes were neglected in the glmertree.

A way of showing how much (or little) all of the predictors contribute to the model is an RF analysis. As the name already indicates, an RF (Breiman, 2001) is a collection of classification trees, involving two different layers of randomness. First, for each tree in the RF, a random subset of the original data set is sampled (bootstrapping), thus creating a variety of subsets of the data for each tree. Second, at each node of the decision tree, rather than considering all features to split on, a random subset of predictors is selected, thus randomly avoiding possibly high-performing predictors. In that way, RFs can handle correlations in the data and help prevent overfitting.

RFs were fit on the joint IndE and SLE data as well as on the SLE data alone. As tree-based approaches were shown to sometimes miss important interactions between predictors (Gries, 2020), a model including interactions of VARIETY with all the other independent variables was fit on the joint data set. In both cases, RFs without interaction predictors performed significantly worse (based on *p*-value) than models that included predictor interactions. The two interaction RFs showed that top-ranking interaction predictors in both varieties are often a combination of sociobiographic and



**FIGURE 6** Variable importance plot for FUNCTION in SLE.

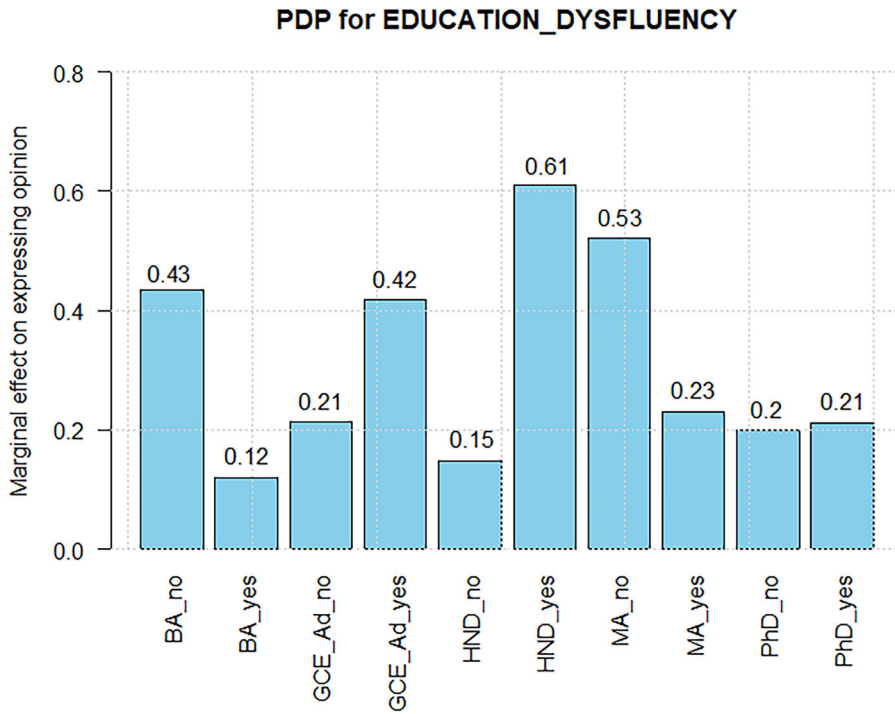
structural factors. As this study is specifically interested in the importance of additional sociobiographic factors, the following investigations will concentrate only on the interaction RF of SLE.

Figure 6 shows the variable importance plot of the RF fitted on the SLE data. The RF has a classification accuracy of 63.06%, which is significantly ( $p = 0.0005$ ) better than a baseline model (53.16%). The plot indicates how much worse the model would get if one was to leave out top-ranking predictors such as the interaction of EDUCATION with DYSFLUENCY, EDUCATION with HOME\_LANGUAGE(s) or ADDITIONAL\_LANGUAGE(s) with DYSFLUENCY.

However, to gain an understanding of the driving forces within said interaction predictors, a closer look at partial dependence plots (PDP) is necessary. Figures 7 and 8 show the marginal effect of the top two interaction predictors on the more frequently occurring level of the response variable (again: expressing opinion) while keeping all other predictors constant.

In Figure 7, one can see that the inclusion of dysfluencies makes no difference in the likelihood of opting for expressing opinion amongst speakers who hold a PhD. All other speaker groups show different behaviour depending on whether dysfluencies are involved in the utterance: Speakers who hold a bachelor's or master's degree tend to express opinion when no dysfluency marker is employed, whereas speakers with a GCE Advanced degree or an HND do not express opinion as readily when no dysfluency marker is involved (hence, they prefer to express opinion with the use of dysfluency markers).

In Figure 8, it first needs to be stated that no speaker with an HND has a home language other than English or another Indo-European language. Therefore, distinctions between speakers with Indo-European versus Dravidian home languages cannot be made for HND holders. However, the PDP shows that for all other degrees, speakers



**FIGURE 7** Partial dependence plot for the interaction of EDUCATION with DYSFLUENCY. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

with home language English or English and another Indo-European language are more likely to use comment clauses to express opinion than speakers with Dravidian home languages. The corpus examples below show that the former group of speakers are more straightforward when expressing opinion, but they also include a variety of other mitigators such as *a bit*, especially when followed by negatively connotated words (examples (8) and (9)).

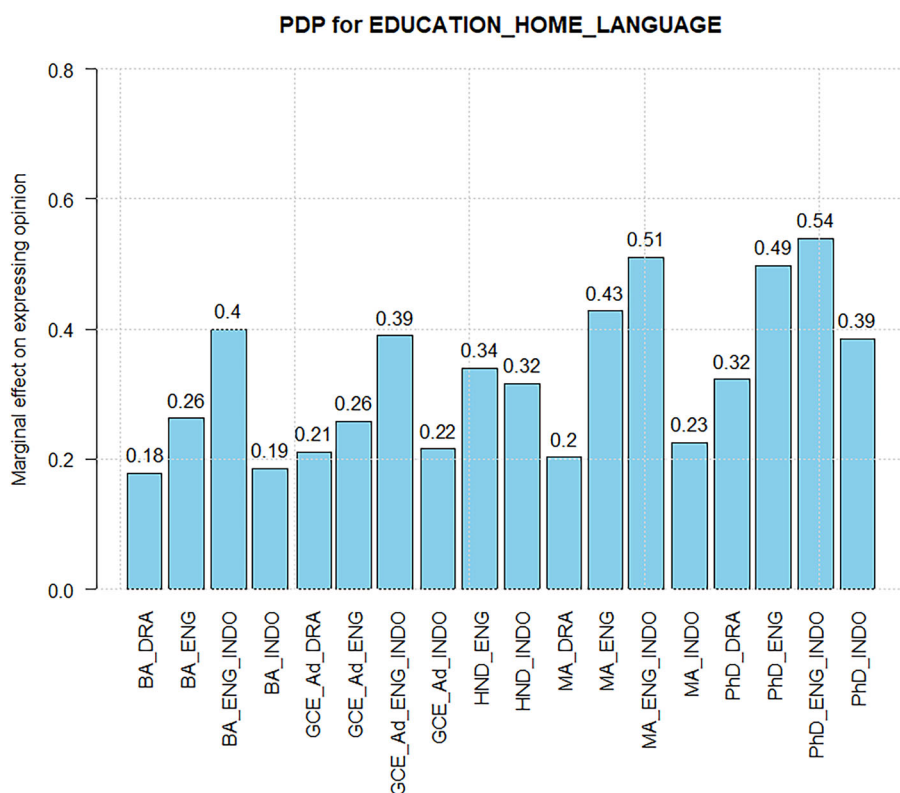
- (8) I **think** generally fights are a bit stupid (ICE-SL:S1A-021#174:1:A)  
 (9) I **think** it was they <„> they didn't say anything but you know no how these YouTube clips I mean half of them don't even describe what exactly is going on ... (ICE-SL:S1A-025#146:1:A)

In contrast to that, speakers with Dravidian home languages feature some dysfluencies and general hesitation phenomena, but no combination of comment clause *I think* with additional mitigating devices (examples (10) and (11)).

- (10) Yeah she she didn't inv she didn't invite me because uh uh I **think** there are some political issues (ICE-SL:S1A-061#243:1:B)  
 (11) But I **think** uh the they are uh perfect ah they are okay  $\leq_{\mu} \geq$  ... (ICE-SL:S1A-001#255:1:C)

## 5 | DISCUSSION

The aim of this study was to provide a first step towards the description of parentheticals in IndE and SLE. This section discusses the study's results in light of the research questions. With regard to the first research question—whether



**FIGURE 8** Partial dependence plot for the interaction of EDUCATION with HOME\_LANGUAGE in SLE. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

or not there are quantitative differences between the two varieties in terms of (preference for) comment clause function—it can be stated that while both varieties employ an overall similar amount of comment clauses, there are differences with regard to the frequencies with which speakers in the two varieties use the two comment clause functions distinguished in the present study. While IndE speakers opt for expressing opinion over 80% of the time, speakers of SLE show no such real preference for either comment clause function. Moreover, the present study supports previous studies in that *I think* is the most frequently used lexical type of comment clauses in IndE and SLE, too, and it is likely that the overall patterns this study showed are influenced mostly by instances of *I think*.

The second research question was concerned with the influence of structural, contextual and sociobiographic factors and which of these (interaction) predictors possibly determine pragmatic speaker choices best in each of the varieties. The analyses above showed that comment clause function to either express opinion or insufficient knowledge is heavily influenced (a) by structural and sociobiographic factors on their own and especially (b) by a combination of structural with sociobiographic factors. The application of a glmertree to the data indicated differences in the speakers' preference of comment clause FUNCTION according to two different independent variables: FORM and SETTING. While IndE speakers behave differently depending on the FORM of comment clause they employ in that they opt for the less frequently occurring level of function (that is to express insufficient knowledge) only when the comment clause takes on the form of *I suppose*, SLE speakers prefer to express insufficient knowledge specifically when communicating in private settings, but they rather opt for expressing opinion in public settings. The IndE's comparably more frequent use of expressing insufficient knowledge when *I suppose* is chosen might coincide with possible patterns of equivalent constructions in indigenous languages available to the speakers, but this assumption would need further investigation. With regard to the SLE preferences, following the ICE corpus design, public settings include classroom interactions or

courtroom discussions. Therefore, on the one hand, the preference to mitigate expressing opinion via the use of a comment clause comes as no surprise as speakers often soften their opinion to avoid a risk of face loss, possibly especially in public and surrounded by speakers who are less familiar. On the other hand, the preference for expressing opinion in public settings is counter-intuitive when considering the fact that, for instance in courtroom discussions, speakers might more often be expected to express insufficient knowledge on a topic than their opinion. However, to see if the majority of cases in which insufficient knowledge is expressed actually stems from courtroom discussions, one would need a more detailed distinction between the various settings summarised in the public category of SETTING.

With regard to the glmertree analysis of SLE only, a tentative qualitative difference between older speakers holding a GCE Advanced degree and speakers with other educational backgrounds shows that *I think* may fulfil the surface function of expressing opinion, but below this surface lie further (rather negative) functions of *I think* that have not yet been captured in current more fine-grained comment clause taxonomies.

Last but not least, the third research question—whether the inclusion of additional sociobiographic factors helps to further predict and explain pragmatic speaker choices—can be answered affirmatively. The RF analysis conducted on the data, specifically the SLE component, showed that additional sociobiographic factors contribute greatly to the classification accuracy of the RF analysis, especially in interaction with other predictors. Generally, none of the predictors is important on their own for the choice of comment clause FUNCTION, which votes in favour of directly coding interactions into the statistical model. Moreover, with EDUCATION, ADDITIONAL\_LANGUAGE(s) and HOME\_LANGUAGE(s) being three of the variables that form top-ranking interaction predictors, it seems more than worthwhile to opt for the inclusion of additional sociobiographic factors, if possible. In addition to that, and in line with previous studies, the RF analysis showed that some of the structural predictors, for example, SCOPE and PLACE of comment clause as well as DYSFLUENCY, proved to be of significance in combination with sociobiographic predictors. In fact, these interactions between sociobiographic and structural factors are amongst the most important predictors for the choice of comment clause function. While there are no qualitative differences in the actual comment clause realisation patterns of the various speaker groups for the interaction of EDUCATION and DYSFLUENCY (meaning that if dysfluencies are used, they mostly reflect in filled and unfilled pauses), one could suggest that university degree holders might be more used to expressing their opinion due to their frequent engagements in academic contexts that call for opinion-making and -expressing, so that dysfluencies are employed less often because those speakers are more confident in expressing opinion. In a similar vein, with the interaction of EDUCATION and HOME\_LANGUAGE(s), one possible explanation for the apparent difference between speakers with Indo-European home languages and speakers with Dravidian home languages in terms of employment of additional mitigating strategies is that, maybe, the use of the equivalent of *I think* in Dravidian languages is not prototypically combined with additional mitigating devices, so that speakers do also not employ such a strategy in English. However, this is only a guess that would need to be verified with actual speakers of Dravidian languages. It would also require further statistical testing to investigate if this is actually a salient pattern distinguishing speakers with Indo-European home languages from speakers with Dravidian home languages.

## 6 | CONCLUSION

This paper has introduced the first multifactorial study on comment clauses in South Asian Englishes. Its aim was to investigate possible patterns of pragmatic speaker choices in IndE and SLE comment clauses and to prove the value of including additional sociobiographic speaker factors. In conclusion, there seems to be a significant difference between IndE and SLE, with IndE speaker choices being influenced by different structural and contextual variables. A qualitative look at the data shows some patterns, but these patterns do not seem unique or variety-specific. Nevertheless, in order to make valid statements about variety-specificity or possible unique South Asian English patterns, future research would have to include BrE data for comparison as well as some information on comment clause patterns of other first languages in India and Sri Lanka in order to make well-founded suggestions about indigenous language influences on pragmatic choices.

Regarding SLE only, this study included two additional sociobiographic factors: the speaker's EDUCATION and OCCUPATION. While OCCUPATION shows to be important only in interaction with other sociobiographic factors (such as ADDITIONAL\_LANGUAGE(S)) and structural factors (such as FORM), EDUCATION is one of the most important independent variables for the choice of comment clause FUNCTION, on its own as well as in interaction with other predictors. A qualitative analysis, again, showed some possible patterns—for instance, regarding the joint influence of EDUCATION and HOME\_LANGUAGE(S)—yet whether these are variety-specific preferences for SLE would need to be cross-checked in future research once respective data are available for varieties like BrE and IndE. In addition, possible underlying functions of *I think* were detected in the present data set, so that future studies could include a more elaborate distinction of various comment clause functions based on initial qualitative investigations of the respective data.

In sum, this paper has highlighted the importance of (a) including a variety of predictor variables (that is structural, sociobiographic and contextual) in general as well as (b) including additional sociobiographic speaker variables if available. By employing multifactorial models, this study has shown that there seem to be variety-specific preferences concerning the choice of the function of the comment clauses investigated in this study. In the future, studies as the present would benefit from more specialised corpora of world Englishes, specifically Outer Circle varieties (Kachru, 1992), that include pragmatic annotation and speakers with a wider socio-economic background. Many of the speakers in the present data set indicate to have English as a home language, possibly even as their first language. Thus, it can be assumed that some of the speakers in this study belong to a small minority of speakers of English in India and Sri Lanka, most likely reflecting a highly acrolectal use of English (which also shows in the interaction of home languages with a higher educational background). Such corpora would allow for future investigations of factors essential to the study of pragmatic speaker choices, such as the degree of imposition or social distance between speakers, as well as language realisations in less acrolectal speech. While this may be a challenging endeavour, it would greatly contribute to our understanding of pragmatic speaker choices, and thus interpersonal communication, in varieties of English around the world.

## ACKNOWLEDGEMENTS

Open access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

There is no conflict of interest to declare.

## ORCID

Julia Degenhardt  <https://orcid.org/0000-0003-0060-1802>

## REFERENCES

- Anthony, L. (2018). *AntConc* (Version 3.5.8). Computer software. Waseda University.
- Arndt, H. (1987). Cognitive verbs and the indication of utterance function. In J. Monaghan (Ed.), *Grammar in the construction of texts* (pp. 24–36). Frances Pinter.
- Bernaish, T. (2015). *The lexis and lexicogrammar of Sri Lankan English*. John Benjamins Publishing Company.
- Bernaish, T. (2022). Comparing generalised linear mixed-effects models, generalised linear mixed-effects model trees and random forests: Filled and unfilled pauses in varieties of English. In O. Schützler & J. Schlüter (Eds.), *Data and methods in corpus linguistics: Comparative approaches* (pp. 163–193). Cambridge University Press. <https://doi.org/10.1017/9781108589314.007>
- Bernaish, T., Mendis, D., & Mukherjee, J. (2019). *Manual to the international corpus of English—Sri Lanka*. Justus Liebig University Giessen, Department of English.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Brinton, L. J. (2008). *The comment clause in English. Syntactic origins and pragmatic development*. Cambridge University Press.
- de Vries, M. (2012). Parenthetical main clauses—Or not? On appositives and quasi-relatives. In L. Aelbrecht, L. Haegeman, & R. Nye (Eds.), *Main clause phenomena: New horizons* (pp. 177–201). John Benjamins Publishing Company.

- Degenhardt, J. (2021). Requests in Indian and Sri Lankan English. *Word Englishes*, <https://doi.org/10.1111/weng.12573>
- Degenhardt, J., & Bernaisch, T. (2022). Apologies in South Asian varieties of English: A corpus-based study on Indian and Sri Lankan English. *Corpus Pragmatics*, 6, 201–223. <https://doi.org/10.1007/s41701-022-00117-8>
- Dehé, N. (2014). *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge University Press.
- Fokkema, M., Edbrooke-Childs, J., & Wolpert, M. (2021). Generalized linear mixed-model (GLMM) trees: A flexible decision-tree method for multilevel and longitudinal data. *Psychotherapy Research*, 31(3), 329–341. <https://doi.org/10.1080/10503307.2020.1785037>
- Fokkema, M., Smits, N., Zeileis, A., & Kelderman, H. (2018). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behaviour Research Methods*, 50, 2016–2034. <https://doi.org/10.3758/s13428-017-0971-x>
- Fraser, B. (1980). Conversational mitigation. *Journal of Pragmatics*, 4(4), 341–350.
- Funke, N. (2022). Pragmatic nativisation of thanking in South Asian Englishes. *World Englishes*, 41(2), 136–150. <https://doi.org/10.1111/weng.12517>
- Funke, N., & Bernaisch, T. (2022). Intensifying and downtoning in South Asian Englishes. *English World-Wide*, 43(1), 33–65. <https://doi.org/10.1075/eww.21064.fun>
- Götz, S. (2013). *Fluency in native and nonnative English speech*. John Benjamins Publishing Company.
- Gries, S. Th. (2018). On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies*, 1(2), 276–308. <https://doi.org/10.1075/jsls.00005.gri>
- Gries, S. Th. (2020). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*, 16(3), 617–647. <https://doi.org/10.1515/cllt-2018-0078>
- Gries, S. Th., & Bernaisch, T. (2016). Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide*, 37(1), 1–25. <https://doi.org/10.1075/EWW.37.1.01GRI>
- Heller, B. (2017). *Stability and fluidity in syntactic variation world-wide: The genitive alternation across varieties of English* [Doctoral dissertation, KU Leuven].
- Kachru, B. (1992). *The other tongue: English across cultures*. University of Illinois Press.
- Kaltenböck, G. (2010). Pragmatic functions of parenthetical I think. In G. Kaltenböck, W. Mihatsch, & S. Schneider (Eds.), *New approaches to hedging* (pp. 237–266). Emerald Group Publishing Limited.
- Kluck, M. (2011). *Sentence amalgamation*. LOT, Netherlands Graduate School of Linguistics.
- Kraaz, M., & Bernaisch, T. (2022). Backchannels and the pragmatics of South Asian Englishes. *World Englishes*, 41(2), 224–243. <https://doi.org/10.1111/weng.12522>
- Lange, C. (2012). *The syntax of spoken Indian English*. John Benjamins Publishing Company. <https://doi.org/10.1075/veaw.g45>
- Lange, C., & Leuckert, S. (2020). Tag questions and gender in Indian English. In T. Bernaisch (Ed.), *Gender in world Englishes* (pp. 69–93). Cambridge University Press. <https://doi.org/10.1017/9781108696739.004>
- Mindt, I. (2003). Is I think a discourse marker? In E. Mengel, H.-J. Schmid, & M. Steppat (Eds.), *Proceedings Anglistentag 2002 Bayreuth* (pp. 473–483). Wissenschaftlicher Verlag Trier.
- Mittwoch, A. (1979). Final parentheticals with English questions—Their illocutionary function and grammar. *Journal of Pragmatics*, 3(5), 401–412.
- Mukherjee, J. (2008). Sri Lankan English: Evolutionary status and epicentral influence from Indian English. In K. Stierstorfer (Ed.), *Proceedings Anglistentag 2007 Münster* (pp. 359–368). Wissenschaftlicher Verlag Trier.
- Persson, G. (1993). Think in a panchronic perspective. *Studia Neophilologica*, 65, 3–18.
- Revis, M., & Bernaisch, T. (2020). The pragmatic nativisation of pauses in Asian Englishes. *World Englishes*, 39(1), 135–153. <https://doi.org/10.1111/weng.12387>
- Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.
- Shastri, S. V. (2002). *Overview of the Indian component of the International Corpus of English (ICE-India)*.
- Yule, G. (1996). *Pragmatics*. Oxford University Press.

**How to cite this article:** Degenhardt, J. (2024). Parentheticals in spoken Indian and Sri Lankan English. *World Englishes*, 1–18. <https://doi.org/10.1111/weng.12696>