Aspect in Learner Writing:

A Corpus-Based Comparison of Advanced Bulgarian and German

Learners' Written English

Inaugural-Dissertation

zur

Erlangung des Doktorgrades der Philosophie

des Fachbereiches 05

der Justus-Liebig-Universität Giessen

vorgelegt von
Svetlomira Rogatcheva
aus Sofia

2014

Gutachter:
1. Gutachter: Prof. Dr. Magnus Huber
2. Gutachter: Prof. Dr. Joybrato Mukherjee
Tag der Disputation:

Table of Contents

PART I

1. Introduction

The study of aspect has been likened to a dark and savage forest full of "obstacles, pitfalls and mazes which have trapped most of those who have ventured into this much explored but poorly mapped territory" (Binnick 1991: 135).

1.1 Aim and motivation of the study

Research on aspect has a very long tradition and dates back to Aristotle, who was one of the first to address the notion of 'aspect' as a feature of the verb which realises a special form of temporality in a language (cf. Klein 1994: 14). Two millennia later, researchers still disagree on what counts as verbal aspect and Comrie's definition of aspect as "the different ways of viewing the internal temporal constituency of a situation" (Comrie 1976: 3) is just one of many definitions, albeit the most quoted one. And yet, what researchers of the English language agree on is the fact that aspect belongs to one of the core areas of English grammar, and its mastery in English as a Foreign Language (EFL) is regarded as the "sine qua non of the mastery of English" (Lorenz 2002: 132), since it often remains "unmastered", even at an advanced level (cf. Meunier and Littre 2013).

The two aspect forms in English most researchers agree on – the progressive and the perfect – have thus been acknowledged as notoriously difficult for learners of English from a wide range of mother-tongue backgrounds and proficiency levels (e.g. Swan and Smith 1987). Explanations for the problematic acquisition and use of the progressive and the perfect range from their inadequate representations in classroom teaching materials (e.g. Römer 2005) to typological differences between the English aspect system and the aspect systems of the learners' native languages. Research on aspect based on large computerised datasets consisting of samples of authentic language – language corpora – has been scarce so far: a small number of corpus-based studies investigate the use of tense-aspect forms in EFL learners' writing, adopting mainly a numerical, quantitative approach (cf. Granger 1996; Virtanen 1997; Granger 1999; Axelsson and Hahn 2001; Housen 2002a; 2002b; Lenko-Szymanska 2007, Davydova 2011 etc.).

The goal of the present study is to fill this gap: I seek to examine how aspect is used in the writing produced by advanced EFL learners from two very different mother-tongue backgrounds – Bulgarian (which has not been investigated so far) and German – from a quantitative and a qualitative perspective. The focus of the study is threefold:

1) to compare and contrast aspect use in advanced Bulgarian and German EFL learners' writing and to answer the question "Do Bulgarian and German EFL learners use English aspect in a targetlike manner and to what extent?"
2) to examine the role of the native language influence on the use of aspect in EFL writing in relation to other learner- and learning-related variables (e.g. proficiency level, exposure to English);
3) to establish areas of common difficulties for both learner populations

Such a contrastive "mastery check" on aspect use in advanced EFL writing is a new area of research at the interface between Corpus Linguistics and Second Language Acquisition worth delving into, since it combines the "best of two worlds": the quantitative rigor of more recent computer-based corpus linguistic approaches with the qualitative methods and insights decades of research into second language use have to offer. The research is based on learner data from the Bulgarian and German components of the *International Corpus of Learner English* (ICLE) (cf. Granger 2009), a collection containing argumentative essays written by advanced learners of English, as well as two sets of native-speaker control data, comprising British and American English writing.

1.2. Terms and Concepts

Since the ongoing debate on the exact nature of verbal aspect and its realisations in English will be elaborated on in detail in the next chapter, this section briefly explains some of the terms and concepts central to the present study. **Learner writing** refers to the written production of learners of English for whom English is a foreign language – i.e. speakers for whom English is not a first language and who do not live in an English-speaking environment (e.g. in Germany and Bulgaria, cf. also Davydova 2011: 8). "Foreign language" will thus be used in the sense of a language "not widely used in the learners' immediate social context [but] which might be [either] used for [...] cross-cultural communication situations" (Saville-Troike 2007: 4) or in the classroom – just not on a daily basis. The terms "second language"

(L2) and "Second Language Acquisition" (SLA) will be used as overarching cover terms for the acquisition of any language "subsequent to learning [the] first language" (Saville-Troike 2007: 2), including the learning processes both within a formal environment (e.g. school or university) and informal (e.g. naturalistic) contexts, alongside with other environments such as immigrant communities in a target-language country (e.g. the UK), or within non-target language environments such as the foreign-language classroom (e.g. Germany and Bulgaria).

In the case of the Bulgarian and German learners of English in the present study, the terms second language (L2) and second language acquisition (SLA) will be used somewhat indiscriminately to refer to the acquisition of English as a Foreign Language (EFL) in a non-English-speaking environment in Germany and Bulgaria. Learner language will also be interchangeably referred to as "Interlanguage" (IL) after Selinker's definition of the intermediate linguistic system "based on observable output" second language learners develop as they move from their first or native language (NL or L1, i.e. German or Bulgarian) towards the target language (TL, i.e. English) (Selinker 1972: 2013-214). The term "advanced learners" will refer to young adults who are English majors in at least their second year of English studies at university (see chapter 5, cf. also Granger 2007: 172).

The corpus-based comparison of advanced EFL learners' written English refers to a new research development at the interface between Corpus Linguistics (broadly defined as the study of language based on authentic language samples, e.g. McEnery and Wilson 1996: 1, see also chapter 4) and Contrastive Linguistics (broadly defined as the subdiscipline of Linguistics which is concerned with the comparison between two or more languages, cf. Lado 1957; Granger 2003 etc., see also chapters 2 and 3) which employs a contrastive analysis of comparable computerised collections of authentic learner and native data in one and the same language or in varieties of this language, i.e. German learner English and Bulgarian learner English vs. British English and American English (cf. Granger 2003: 19-20). The corpus-based comparison of learner language has the advantage of confronting existing (and possibly dated) theories about SLA with actual, authentic data – both learner and native speaker data, and thus either confirm or challenge learner language descriptions.

1.3. Preview

The study is organised in ten chapters: chapters 1 to 4 (Part I) offer a comprehensive discussion of the theoretical framework adopted in the present study, whereas chapters 5 to 10 provide the empirical part (Part II). Chapter 2 introduces aspect as a formal and semantic category and elaborates on the debate about aspect and aspectuality in English, its formal realisations, as well as quantitative distribution and meaning variation. The final section of the chapter deals with a contrastive comparison of the aspect system in English and the aspect systems of Bulgarian and German. Chapter 3 reviews the major SLA theories on the acquisition of aspect – focussing on English as a target language – and summarises the existing approaches and hypotheses about general trends in L2 aspect development (e.g. Aspect and Discourse Hypotheses). Further, the chapter investigates the role of L1 on the acquisition and use of English aspect. Chapter 4 turns to the latest corpus-based developments in the study of grammar and reviews previous research on L2 aspect use based on learner corpus data. Chapter 5 introduces the research methodology behind the quantitative and qualitative part of the analysis, as well as the software tools and corpus data. The quantitative analysis of the frequency distribution of aspect forms in learner and native writing is presented in chapter 6, while chapter 7 deals with the lexicogrammatical variation in the use of the progressive and the perfect such as the distribution of lexical verb types across progressive and perfect aspect markers, the distribution of progressive and perfect verb phrases across main and subordinate clauses, and the co-occurrence of aspect forms with temporal adverbials and contracted auxiliaries. Chapter 8 discusses Bulgarian and German EFL learners' misuse of aspect forms from a qualitative and a quantitative perspective and proposes a new, corpus-based method of measuring learners' targetlike use. A comprehensive discussion of the results with respect to the interplay between different factors influencing aspect use in advanced EFL writing such as transfer from L1 or L2 exposure is offered in chapter 9, which also puts forward a model for the analysis of L2 aspect use, together with a synthesis of the results. Finally, a brief conclusion and an outlook for future research on aspect use in advanced EFL learners' writing is given in chapter 10.

2. Aspect in English: Theoretical, Quantitative and Contrastive Perspectives

There are many different ways in which human languages can express physical time and aspect is just one of them.[1] Every human language has its own ways of locating events and situations in time; however, the forms, shapes and the extent to which temporal distinctions are encoded through lexical and grammatical means in a language can differ dramatically from one language to the next. Research on temporality and the formal means of temporal realisation in natural languages has a very long tradition and dates back to Aristotle, who was one of the first philosophers to characterise verbs (*rhemata*) in relation to past, present or future time and to address the notion of 'aspect' as a feature of verbs which realises a special form of temporality (Klein 1994: 14). Similar to the perception of space, "[t]he experience of time is fundamental to human cognition and action" (Klein 2009: 39 and has been recognised and analysed as such by western philosophy throughout the centuries. Consequently, research on the expressions of temporality in the world's languages has been mostly guided by western thought; furthermore, the temporal concepts originating from this research are largely oriented towards Indo-European languages such as Greek, Latin, English or German, all of which have an obligatory formal encoding of temporal categories (cf. Klein 2009: 39-41).

Within the Western research tradition on temporality, two major approaches have been adopted to characterise the expressions of time in language: a meaning-oriented approach, concerned with the different means the world's languages employ to express the concept of temporality linguistically, and a form-oriented approach, which starts with an analysis of the language-specific formal categories before characterising their meanings. The meaning-oriented approach has been employed by various typological studies which explore the boundless possibilities available in the world's languages to express temporal concepts, ranging from grammatical and lexical means to pragmatic and discourse devices (e.g. Comrie 1976; Bybee et al. 1994; Bybee 1985; Dahl 2000 etc.). The form-oriented approach takes on a more narrow perspective by investigating language-specific forms which carry temporal meanings in a particular language or languages, comparing them cross-linguistically and then focussing on the differences in form and meaning.

---

[1] Although temporality plays a marginal role for the present study, it contributes to the understanding of aspect from a holistic perspective; hence a brief word on temporality is in order here

The present study is an investigation of aspect in two corpora of written learner English and as such, it will focus on the formal expressions of aspect in learner English by comparing them to aspect forms in native-speaker written English. Therefore, the study will adopt a largely form-oriented perspective to aspect in English as a second language, whereby discussions of meaning deviations of aspect forms in learner writing from native-speaker writing will inevitably follow an initial analysis of aspect forms. The goal of the present chapter is to introduce aspect in English as a formal and semantic category from a theoretical, empirical, and a contrastive perspective. The chapter will start with a theoretical discussion of the category of aspect and aspectuality in general and aspect in English in particular, and will delineate the category of aspect from other categories such as tense and actionality (section 2.1); furthermore, it will proceed with a more detailed description of the English aspect forms, their quantitative distribution and meaning variation in contemporary written and spoken British and American English (section 2.2). Finally, a brief contrastive comparison between aspect in English and the formal realisation of aspectuality in German and Bulgarian as native languages of the two learner populations in the present study will also be given (sections 2.3. and 2.4.), including some hypotheses concerning the possible difficulties Bulgarian and German learners of English might experience when acquiring and using English aspect.

2.1. Approaches to Aspect

Albeit ancient as a concept, the term "aspect" originated from the study of Slavic grammar in the early 19th century as a direct translation from the Russian word *vid* (literally 'view' or 'type') (cf. Gonda 1962: 9 in Comrie 1976: 1; Brinton 1988: 2; Binnick 1991: 135; Tobin 1993: 3). The first mention of the term "aspect" in English appeared as late as 1853 in the Oxford English Dictionary (OED) (Binnick 1991: 139-140), where it was introduced from Slavic linguistics to denote the typically Slavic opposition between perfective verbs (those expressing a complete action, as in e.g. the Russian *napisat* ('write' as in 'write (up) a novel', perfective) – and imperfective verbs (those expressing an ongoing action, e.g. Russian *pisat* ('is/was writing', 'used to write', imperfective) (cf. Comrie 1976: 3). Ever since its first appearance in the OED, there has been an ongoing controversy in the research on temporality concerning the exact nature and definition of aspect – resulting into two "fundamental approaches to verb 'aspect' – a 'temporal' approach and a 'non-temporal' approach" (Tobin 1993: 5). Tobin (1993: 5) summarises both approaches by offering a collection of definitions following the 'temporal' and the 'non-temporal' view. According to the temporal approach,

aspect has been commonly defined in a number of different ways, ten of which are presented below (Tobin 1993: 5):

> (i) 'a way of conceiving the passage of time' (Friedrich 1974: 2)
> (ii) 'the manner in which the action of the verb proceeds [presumably in time] (Gonda 1962: 12)
> (iii) 'the way in which a process takes place in time or is placed in time' (Gonda 1962: 10)
> (iv) '[the] temporal values inherent in the activity or state itself' (Jakobson 1971: 130 – 147)
> (v) '[signifying] the relative duration or punctuality along a time line that may inhere words or constructions' (Friedrich 1974: 1)
> (vi) 'the name for the function of discriminating the kinds of temporal "things" which may be (linguistically) "located" in the sequential order of time' (Taylor 1977: 164 – 5)
> (vii) 'reference to one of the temporally distinct phases of the evolution of an event through time' (Johnson 1981: 152)
> (viii) 'the TEMPORAL QUALITY OR CONDITION of an event with respect to itself, in terms of such things as inception, repetition, duration, punctuality, etc.' (Freed 1979: 10, original emphasis)
> (ix) 'the expression of 'the moments or stages of the process' (Gonda 1962: 11 on Rasmussen)
> (x) 'different ways of viewing the internal temporal constituency of a situation' (Comrie 1976: 3)

Among these ten different definitions, Comrie's definition of aspect has been one of the most widely quoted and will be reviewed in more detail in the next section. The 'non-temporal' approach summarised by Tobin (1993: 5 – 6) includes definitions such as:

> (i) 'whether the speaker looks upon an action or event in its entirety, or with special reference to some part' (chiefly the beginning or end) (Kruisinga 1931: 221)
> (vi) a 'system of orientation' (Tobin 1986, 1989) […] based on spatio-temporal-existential boundaries conceptualized by the encoder at the here-and-now point of encoding';
> (vii) 'a speaker's viewpoint or perspective on a situation…as either completed (perfective aspect), or as ongoing (imperfective aspect), or repeating (iterative or habitual aspect)' (Brinton 1988: 3)

To a certain extent, the two approaches quoted above use overlapping defining terms such as 'viewpoint', 'viewing', 'phases' or 'stages' and thus confirm that the two seemingly contradictory research strands lack clear-cut boundaries. Tobin (1993: 4-5) notes that an alternative unified approach to aspect has also been attempted by researchers in the 1980s in what is referred to as the tense-aspect-modality (T-A-M) theory (cf. Hopper 1982 in Tobin 1993: 5). Still, most traditional approaches to aspect stick to the two-tier differentiation between aspect as "the manner and way in which the action of the verb proceeds" (Gonda 1962: 12-13), and the second type of aspect which "express[es] the moments or stages of the process" (Gonda 1962: 11). The former definition has commonly served as the basis for a term variably called "actionality", "situation aspect", "kind of action" (*Aktionsart*), whereas the latter has been called "viewpoint aspect" or aspect 'proper' (cf. Smith 1983). Smith

characterises the situation type of aspect (*Aktionsart*) according to the four categories "distinguished by Aristotle and others [such] as Activity, Achievement, Accomplishment and State" in opposition to the second type of aspect – the "*aspectual perspective* – simple or progressive in English", which renders the speaker's perspective on the situation (Smith 1986: 97). These four categories will be further discussed and exemplified in the following section, which will focus on the opposition between viewpoint aspect and situation aspect in greater detail.

2.1.1. Situation Aspect vs. Viewpoint Aspect

As mentioned in the previous section, the term 'aspect' is a "much-discussed but still rather elusive category" (Tobin 1993: 3), and in order to grasp it fully, it is important to draw a more fine-grained differentiation between the two major notions (or groups of notions) which fall under the category of aspect. The traditional classification differentiates between viewpoint aspect or the type of aspect which is realised by grammatical means "to express various meanings which have to do with how the speaker wants to represent the internal temporal structure of the situation" (Declerck 2006: 28), and situation aspect which "represent[s] the situation as having particular ontological features, such as 'static', 'durative', etc." (Declerck 2006 : 49). Smith distinguishes between "two components of sentential aspect […] SITUATION ASPECT [which] involves type of situation, e.g. event or state; VIEWPOINT ASPECT [which] involves type of perspective, e.g. simple or progressive" (Smith 1983: 479, original emphasis). The latter type of aspect (viewpoint aspect) will be at the heart of the present investigation, whereas the former type of aspect (situation aspect) will be examined only in relation with viewpoint aspect from a theoretical (chapter 3) and a methodological (chapters 5 and 7) perspective. The opposition between the two types of aspect represents the traditional two-tiered approach to aspect discussed in the previous section, where viewpoint aspect is realised primarily via grammatical means and situation aspect is coded via lexical means. Although Tobin acknowledges that "the 'grammar' and the 'lexicon' [cannot be] viewed as being separate and distinct" (Tobin 1993: 7), thus pleading for a unified, "more isomorphic, semiotic or sign-oriented" (ibid.: 7) approach, this differentiation is still necessary when investigating the relationship between the two types of aspect in actual language use, and in particular in non-native language use. Therefore, the present section will focus on this traditional distinction, starting with situation aspect and providing examples for the various types of situation aspect and their ontological features.

Situation aspect has received various names in the literature on aspectual research, the most common of which "inherent lexical aspect", "lexical aspect" (e.g. Andersen and Shirai 1996: 530), "ontological aspect", and lastly, *Aktionsart* (cf. Binnick 1991: 144; Declerck 2006: 49).[2] Situation aspect refers to the inherent characteristics of the verb phrase which are independent of the fact whether the verb phrase is marked for grammatical aspect and the temporal context in which the verb phrase is used (cf. Declerck 2006: 49). These inherent features belong not only to the verb itself, but also to other elements of the verb phrase such as the objects or complements of the verb – therefore, the inherent lexical aspect of a verb is a property of the whole verb phrase, rather than of the verb in isolation (cf. Klein 1994: 31). To illustrate, there is an obvious difference between the two utterances *he was sitting* and *he was sitting down* – in the first case there is no change of state (i.e. the person is sitting during the whole situation described), whereas in the second there is a change of state from standing to sitting; therefore, although both utterances are marked for the progressive, they differ in their inherent situation aspect (cf. Klein 1994: 30). As a rule, the situation aspect of verb phrases is realised by lexical means, through the inherent lexical meaning of verbs and other elements of the verb phrase, as well as through derivational morphology (cf. Comrie 1976; Brinton 1988).

> *Aktionsart*, in its narrower sense, relating only to the lexical verb in question, provides lexical information. The arguments and adjuncts of the verb may provide further information, i.e. information coming from the narrowest contexts. Further information may be provided by the broader context, including here both the linguistic context (co-text, in some terminologies) and the context of the situation, or more generally pragmatics. (Comrie 2001: 43)

The inherent lexical aspect of verb phrases has been the object of philosophical and linguistic investigations ever since "Aristotle's distinction between 'verbs of kinesis' and 'verbs of energeia'" (Klein 1994: 31) and has been classified in terms of four basic categories: Activities, Achievements, Accomplishments and States (cf. Smith 1986: 97). This fourfold classification is the most widely-quoted classification of situation aspect which was initially developed by Vendler (1957: 98-99). Referring to English, the four basic categories are defined and exemplified as follows:

(1) **State** – that which has no dynamics and continues without additional effort or energy being applied, e.g. *see, love, hate, want*.
(2) **Activity** – that which has duration, but an arbitrary end point, and is homogeneous in its structure, e.g. *run, sing, play, dance*.

---

[2] Although some researchers claim that there are differences between these four terms, the present study will use them interchangeably to denote situation aspect

(3) **Accomplishment** – that which has some duration, but a single clear endpoint, e.g. *run a mile, make a chair, build a house*.

(4) **Achievement** – that which takes place instantaneously, and is reducible to a single point in time, e.g. *recognize, die, reach the summit*. (adapted from Andersen and Shirai 1996: 531-532)

Although illustrated with English examples, Vendler's fourfold classification is based on cognitive distinctions speakers make irrespective of their native languages, and has been successfully adapted and widely employed by various studies investigating a number of the world's languages, thus qualifying as a "cognitive universal" (Andersen and Shirai 1996: 532). These four categories have also been further specified in terms of three semantic features: punctuality, telicity and dynamicity. Punctuality is opposed to durativity and refers to "the quality of a situation that does not last in time […], one that takes place momentarily" (Comrie 1976: 42). A punctual situation has no duration and is therefore incompatible with imperfectivity or progressivity as a subtype of imperfectivity; to illustrate, *At this point, John is reaching the summit* is inappropriate since *reach the summit* is a punctual situation (Comrie 1976: 42- 43). Telicity (from Greek *télos* 'end', cf. Radden and Dirven 2007: 179) is a semantic quality of verbs and verb phrases which has not received a "uniform treatment" (Kabakčiev 1989: 13), in particular with regard to the differences between the 'Eastern' and 'Western' approach to situation aspect (Dahl 1981 in Kabakčiev 1989: 14). In general, telicity refers to "the inherently conclusive and definitive endpoint of an event" (Radden and Dirven 2007: 179). Telicity is sometimes used interchangeably with 'boundedness' (cf. Kabakčiev 1989: 14), a semantic quality which expresses whether "the content expressed has a 'left' and a 'right' boundary ('unbounded' vs. 'bounded', often contrasted as 'processes' vs. 'events')" (Klein 1994: 31). For the purposes of the present study, the term 'telicity' will be used to differentiate between telic and atelic situations or situations with and without a conclusive endpoint – thus, the difference between the two sentences *John is making a chair* and *John is singing* (examples taken from Comrie 1976: 44) is the conclusive endpoint of the former situation – once the chair is ready, John can no longer keep making it (i.e. *making a chair* is a telic situation), whereas John can keep singing (i.e. *singing* is an atelic situation). Dynamicity is opposed to stativity and "seems reasonably clear intuitively" (Comrie 1976: 48); however, this distinction is not always straightforward, especially in the case of particular verbs (e.g. *stand*). Comrie uses the term 'phase' to differentiate between the two – phase refers to "a situation at any given point of time in its duration" (Comrie 1976: 48) – and exemplifies it with the verb *know*, all phases of which can be characterised as identical and thus stative, as

opposed to the verb *run*, which features different phases and necessarily involves change (cf. Comrie 1976: 49)[3].

With regard to the further categorisation of Vendler's four categories in terms of the three semantic features, it can be assumed that achievements are punctual and telic since they have no duration, but a conclusive endpoint (e.g. *recognise*), whereas accomplishments are durative and telic, i.e. they have some duration (e.g. *build a house*) but a conclusive endpoint. In contrast, states and activities are both atelic and durative (e.g. *live* and *run*); moreover, all inherent lexical aspect verb types apart from states (activities, accomplishments and achievements) are dynamic (see table 2.1).

Semantic Features of the Four Types of Inherent Lexical aspect				
	Lexical aspectual classes		Accomplishments	Achievements
semantic features	States	Activities	(telic events)	(punctual events)
Punctual	-	-	-	+
Telic	-	-	+	+
Dynamic	-	+	+	+

Table 2.1. Semantic features of the four categories of inherent lexical aspect (adapted from Andersen and Shirai 1996: 532)

Vendler's classification has been further developed in subsequent theoretical studies to include two additional subcategories such as punctual activities which are atelic, e.g. *jump*, and punctual states involving "inert perception" (Leech 1971: 27), e.g. *notice*. Leech illustrates these additional semantic properties by arguing that punctual activities in the progressive often have a quality of iteration without an inherent end point, e.g. *She is jumping*, and that in contrast to accomplishments and achievements, which are perfectly acceptable in the progressive, punctual states in the progressive sound awkward and unnatural, e.g. *John is noticing a scratch on the woodwork* (Robison 1995: 350; Leech 1971: 27-29). As an alternative solution to the differentiation between punctual and durative lexical verb types, Robison (1995: 351) proposed a six-fold classification of inherent lexical aspect, including states, activities and events, subdivided into two categories each: durative and punctual. The six-fold classification is summarised in table 2.2:

---

[3] With regard to *stand*, Comrie (1976: 49) notes that although it does not involve change of phase, it may do so, e.g. a book standing on a shelf may change its position if the shelf is moved

	stative	dynamic	
		atelic	telic
durative	durative state	durative activity	durative event
punctual	punctual state	punctual activity	punctual event
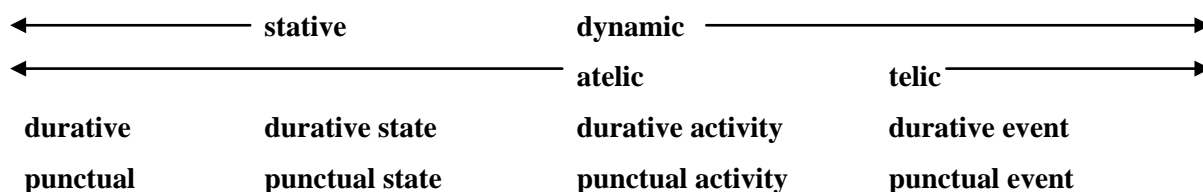
Table 2.2. A six-fold classification of lexical aspect (adapted from Robison 1995: 351)

Despite the continuous efforts to group lexical verbs into clear-cut lexical categories, a strict delineation between Vendler's four types of inherent lexical aspect is not always possible – thus, many lexical verbs can be classified in terms of more than one category. To illustrate, *think* can be both a state and an activity verb (cf. Brinton 2000: 143) – e.g. *I think that he must be in his forties* is thus a state, whereas *She was thinking about its replacement* can be classified as an activity. Therefore, since inherent lexical aspect is a compositional property of verb phrases as whole entities, including their nominal arguments, lexical verbs should never be considered in isolation. Apart from presenting comprehensive lists with the classification of verbs according to their inherent lexical aspect (e.g. Dowty 1979: 66-71, Bardovi-Harlig and Reynolds 1995: 130; Collins 2002: 94 etc.), a number of studies have used and developed several diagnostic tests in order to distinguish between Vendler's four categories (cf. Andersen and Shirai 1995: 749; Brinton 1998: 242; Brinton 2000: 143 – 147). The diagnostic tests employ various methodological procedures, ranging from questions such as *for how long?* or *how long did it take?*, to conditions such as *If X Ved in Y time, then X was Ving during that time* and the possibility of insertion of adverbs such as *almost* (cf. Andersen and Shirai 1995: 749; Brinton 1998: 242). Thus, the diagnostic tests fall into three main categories:

(1) tests of **adverb modification** (testing whether the verb phrase can be combined with a specific adverbial, e.g. *almost*)

(2) tests of **aspect modification** (testing whether a verb phrase like e.g. a stative tolerates the progressive or the so-called *–ing* test)

(3) tests of **presuppositions and implications** (testing whether the verb phrase in question presupposes or implies something else for a specific time subinterval of the main time interval of the verb phrase, e.g. *if X Ved in Y time, then X was Ving during that time*)[4] (cf. Klein 1994: 34).

---

[4] The diagnostic tests for inherent lexical aspect will be illustrated in more detail in chapter 5.

These three types of diagnostic tests have been extensively used to categorise verb phrases, and in particular in relation to verb inflections in first and second language acquisition data in order to examine possible interdependence between the lexical aspect of verbs and children's and second-language learners' use of grammatical markers of temporality (e.g. Klein 1994: 34; Andersen and Shirai 1995; Bardovi-Harlig 2000; 2002 etc.). The relationship between tense-aspect markers and inherent lexical aspect has become the object of extensive research not only in English, but also in a number of other languages where researchers have identified universal patterns of interdependence between the gradual development of grammatical inflections in children's and second-language learners' language and lexical aspect or *Aktionsart*[5].

In contrast to *Aktionsart* which is inherent to the verb phrase and depends on its semantic contents, viewpoint aspect or the so-called aspect 'proper' is a "matter of the speaker's viewpoint or perspective on a situation" and is grammatical (Brinton 1988: 3). Brinton notes that "[t]he differentiation of aspect and [A]ktionsart has, in fact, been approached from a number of different directions: in terms of the contrasts 'grammatical' vs. 'lexical' aspect, 'subjective' vs. 'objective' aspect, 'aspect' vs. 'character'" (Kruisinga 1931: 230 – 7 in Brinton 1988: 3). The major difference between *Aktionsart* and aspect 'proper' according to the majority of the conventional aspect theories lies in the fact that aspect 'proper' is grammatical, since it is realised by grammatical markers of the verb phrase like "verbal inflectional morphology and periphrases" (Brinton 1988: 3), which make it a matter of the speaker's subjective choice of grammatical means, whereas *Aktionsart* is objective, since it is inherent and does not depend on the perspective of the speaker (cf. Brinton 1988: 3). Thus, the two major strands of aspectual research – those dealing with *Aktionsart* and those dealing with aspect 'proper' focus on two different sides of verb forms – "one concentrates on the grammatical meaning of verbal forms, while the other concentrates on the lexical meaning of verbs and their complements" (Brinton 1988: 5).

The most well-known and widely-quoted definition of viewpoint aspect is Comrie's definition, which characterises viewpoint aspect as "the different ways of viewing the internal temporal constituency of a situation" (Comrie 1976: 3). Thus, a situation may either be viewed as a whole, "without necessarily distinguishing any of the internal structure of the

---

[5] The theoretical framework and empirical studies investigating this interdependence will be reviewed in greater detail in chapter 2

situation" (Comrie 1976: 4), as in the English example *John entered the room'*, or as a structure consisting of different internal portions, which can be looked at from within (e.g. by looking at the inner constituency of the action), without a reference to the beginning or the end of a situation, as in e.g. 'John was reading'. In the former example the situation is seen as a completed whole from the outside – i.e. it is perfective, whereas in the latter example the situation is seen as an ongoing process from the inside – i.e. it is imperfective.

With regard to the various definitions of viewpoint aspect, Comrie (1976: 11) and Brinton (1988: 5) note that, similar to the confusion surrounding the troublesome delineation between viewpoint aspect and situation aspect, there has been a multitude of terms used to define the different subcategories of viewpoint aspect. A schematic representation of the 'aspect terminology confusion' summarised by Brinton (1988: 5) is offered below[6].

Category	Terms
Perfective aspect	perfective, aorist, punctual, resultative, momentaneous
Imperfective aspect	imperfective, progressive, imperfect linear, continuative, durative, cursive
Perfect aspect	perfect, perfective
Ingressive aspect	inchoative, ingressive, inceptive
Continuative aspect	Continuative, progressive
Egressive aspect	egressive, resultative, terminative, effective, finitive
Habitual aspect	Iterative, frequentative

Table 2.3. Aspect terminology confusion (adapted from Brinton 1988: 5)

Having discussed the major differences between *Aktionsart* and viewpoint aspect and before turning to the individual forms and realisations of viewpoint aspect in English (and addressing some of the terms illustrated above), the next section will focus on one further important differentiation – the differentiation between viewpoint aspect and tense.

2.1.2. Tense vs. Aspect

Similar to the disagreement concerning the delineation between viewpoint aspect and *Aktionsart*, lexicon and grammar, there is no uniform treatment of the opposition between viewpoint aspect and tense. Most researchers agree that both are categories of the verb phrase

---

[6] A detailed explanation of all aspect terms summarised by Brinton (1988) goes beyond the scope of the present study; besides, not all aspect terms can be adequately illustrated with English examples only – therefore, the present section will not attempt to explain this confusion of terminology and will only try to "scratch the surface" of the aspect terminology debate.

used to express temporal relations, realised exclusively via grammatical means (Huddleston 1984, Quirk and Greenbaum 1990, Radford 1988, Leech and Svartvik 1994, Klein 1994, Aarts 1997, Hahn 2007); in addition, most researchers identify one major difference between the two categories – whereas tense locates a situation with respect to external physical time, usually the moment of speaking and is thus a deictic category (i.e. pointing to a language-external matter), aspect does not relate the time of the situation to any external moment, but only renders the internal temporal make-up of a situation, irrespective of the moment of speaking, which makes it non-deictic (cf. Comrie 1976: 5).

Tense itself has to be distinguished from physical or calendar time, which is an extra-linguistic phenomenon and cannot be captured unless some arbitrary reference points are introduced to measure its lapse, for example specific events in a particular culture such as the birth of Jesus Christ in the Christian world (Gast and König 2009: 79-80), or the founding of Rome in the ancient Roman empire (Comrie 1985: 14). Comrie (1985: 13) argues that:

> If time had a beginning, we do not know where that beginning was, so we cannot locate anything else relative to that beginning (other than, trivially, by saying that the situation is posterior to that beginning).

On the other hand, time can be measured linguistically if a situation is described as happening before, after, or is simultaneous with another arbitrary reference point in time – usually the moment of speaking. The majority of linguistic descriptions of time favour an abstract representation of time which consists of a straight line with an indefinite length with the moment of speaking in its centre illustrating the present moment, the happenings before the moment of speaking located to the left illustrating the past, and the happenings after the moment of speaking located to the right signalling the future:

PAST   **FUTURE**
now (the present moment)
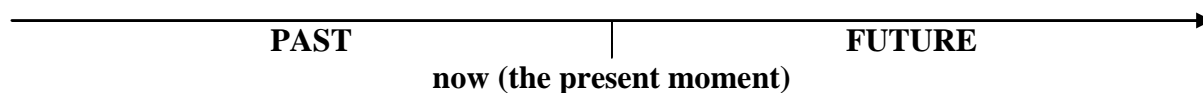
Figure 2.1. Traditional representation of time I (adapted from Quirk 1985: 175)

On the basis of these three intervals, one can loosely refer to the 'present', 'past' and the 'future', which is the traditional system of time conceptualisation utilised in many languages[7]. Klein (1994: 21 – 35) summarises the standard theory of tense by identifying two parameters: **S** for the moment of speaking and **E** for the event being referred to, where **E** can be either a

---

[7] not all languages have grammatical categories like present, past and future to refer to external time, although most languages have lexical means like temporal adverbials (cf. Comrie 1976: 6)

point in time or an interval "occupied by the situation to be located in time" (Comrie 1985: 122). The formal representation of tense according to this standard theory incorporates as a result three different possibilities, as represented by Klein (1994: 21):

is	**E simul S**
was	**E before S**
will be	**E after S**

Table 2.4. Traditional representation of time II (adapted from Klein 1994: 21)

The standard two-parameter representation of tense has been found inadequate by many subsequent studies (e.g. Reichenbach 1952; for a review see Klein 1994; 1995 etc.), all of which have looked for a more adequate representation of complex tense-aspect forms such as e.g. the past perfect in English[8]. These subsequent tense models have tried to overcome the so-called "Aristotelian dilemma" (Klein 1994: 24) or the inadequacy of the standard two-parameter temporal system to account for more complex tense-aspect relationships, mainly by introducing a third parameter in addition to the other two. Reichenbach's (1952: 289-290) tense model is the first and most widely-quoted three-parameter model to introduce a "point of reference" **R** in addition to the point of speech **S** and the event **E**. The point of reference **R** is used to e.g. differentiate between the simple past and the past perfect – in a sentence like "When Mary came to the party, John had left" (Klein 1994: 25) there are two different events in the past – John's leaving and Mary's coming to the party – where the event **E** (John's leaving) precedes the point of reference **R** (Mary's coming to the party) and the point of speech S (now)[9]. Klein (1995: 143) argues in favour of a revised model of Reichenbach's three-parameter temporal system, where tense and aspect can be differentiated with the help of the following parameters: **T-SIT** (referring to the time of the situation, parallel to **E** in the traditional theories), **TU** (the time of the utterance, similar to S) and a third parameter **T-ASS** (time of the assertion), which signals "the time for which an assertion is made by that utterance" (Klein 1995: 143). Whereas the time of the utterance **TU** is normally deictically given (cf. Klein 1995: 142), the time of the assertion **T-ASS** is temporally related to the time of utterance **TU**, so that in a sentence like "Peter was cheerful" (cf. Klein 1995: 142), the time of the assertion **T-ASS** is a "subinterval of the entire situation time, and only for this particular subinterval **T** it is asserted that it precedes the time of utterance". According to

---

[8] the English perfect has been variously defined as a tense (e.g. Reichenbach 1952, Klein 1994, Declerck 2006), an aspect (Comrie (1976), Quirk et al. (1985), Biber et al. (1999) etc. or as a third type of a category (cf. Kortmann 1995) – this discussion will be briefly reviewed in section 2.1.4

[9] Although intuitively easy to grasp, Reichenbach's point of reference R has been severely attacked for its lack of clarity – for a comprehensive review of Reichenbach's (1952) model, its subsequent adaptations and its critics, see also Klein (1994; 1995)

Klein's modified three-parameter theory, tense and aspect can be defined as temporal relations between **TU**, **T-ASS** and **T-SIT**, so that:

> Tense is a temporal relation between TU and T-ASS.
> Aspect is a temporal relation between T-ASS and T-SIT. (Klein 1995: 143)

To illustrate, the time of assertion **T-ASS** for imperfective aspect lies within the time of situation **T-SIT** in e.g. all of the three examples *Eva is sleeping/Eva was sleeping/Eva has been sleeping* which share imperfectivity as a common denominator (and the same relationship between **T-ASS** and **T-SIT**), but a different relationship between the time of utterance (**TU**) and the time of assertion (**T-ASS**) – as in the following examples (Klein 2009:15):

	Perfective	**Imperfective**
Before TU	Eva slept.	Eva was sleeping.
At TU	Eva sleeps.	Eva is sleeping.
After TU	Eva will sleep.	Eva will be sleeping.
Before TU	Eva has slept.	Eva has been sleeping.

With the help of these three parameters and the above relations, the following tenses and aspects in English can be identified following Klein (1995: 144):

TENSE		ASPECT	
FUTURE	TU before T-ASS	IMPERFECTIVE	T-ASS in T-SIT
PRESENT	TU INCL T-ASS	PERFECTIVE	T-ASSOVL[10] T-SIT and TIME afterT-SIT
PAST	TU AFTER T-ASS	PERFECT	T-ASS AFTER T-SIT
		PROSPECTIVE	T-SIT AFTER T-ASS

Table 2.5. Tenses and Aspects in English (adapted from Klein 1995: 144)

Both the traditional two-parameter tense models and Reichenbach's and Klein's three-parameter tense models are "deictic and relational" (Klein 1994: 19), since they relate "entities to a reference point" (Comrie 1985: 14), the most obvious reference point being 'now' or the present moment of speaking (**TU**). In contrast to tense, aspect is non-deictic and non-relational, since it only renders the perspective of the speaker with regard to the situation described – as completed, ongoing, imminent etc., independent of the time "which the event, action, process etc. occupies on the time axis" (Klein 1994: 16), i.e. a pure relationship between **T-ASS** and **T-SIT**. The present study will focus predominantly on this relationship

---

[10] In Klein's terminology, OVL stands for overlapping, i.e. signalling that the two intervals a and b overlap, AFTER stands for a is fully after b, and INCL stands for a is fully included in b)

in native and learner English, although certain combinations between tense and aspect such as e.g. the present perfect or the present progressive in particular will also be investigated in further detail. It is important to mention, however, that Reichenbach's and Klein's parameters do not refer to inflectional categories, but only to "abstract temporal relations" (Klein 1995: 144) and as such will only be of secondary interest for the present corpus-based study, which follows a form-oriented approach to the use of aspect in learner language.

From a formal perspective, Comrie (1976: 5 – 6) (as well as more recent corpus-based grammars and reference works outlined in section 2.2.) identifies two distinct aspect forms in Modern English: a progressive aspect, which in general terms views the action as incomplete; and a perfect aspect, which signifies past time with some current relevance. The following two sections will briefly illustrate the two types of viewpoint aspect in English, together with their general meanings; section 2.1.3. will start with the traditional opposition between ongoing and completed actions, realised by the progressive vs. the simple aspect in English, whereas section 2.1.4. will continue with the debate on the status of the perfect in English as a special type of viewpoint aspect, tense or a third type of category. The remainder of the present chapter (section 2.2.) will turn to a more detailed review of the most recent empirical and quantitative approaches to aspect forms and their meanings in contemporary native English, followed by a short contrastive comparison between aspectual realisations in German and Bulgarian as native languages of the learners in the present study (sections 2.3. and 2.4.).

2.1.3. The Progressive

Most theoretical works agree on the nature of the progressive in English as a viewpoint aspect (cf. Comrie 1976; Bybee et al. 1994 etc.). The English progressive is generally characterised as a category of the verb phrase which expresses an ongoing event or action at a specific reference time (cf. Bybee et al. 1994: 126) or as "a situation in progress" (Comrie 1976: 33) and is formally realised by the auxiliary verb *be* followed by the *–ing* participle (e.g. Biber et al. 1999: 460). The progressive typically combines with dynamic verbs as in e.g. *I am walking*[11] and conveys "actions that require a constant input of energy to be sustained" (Bybee et al. 1994: 126). Comrie (1976: 33) notes that traditional definitions of the progressive do not necessarily delineate it from the category "imperfective", which itself

---

[11] see section 2.2.1.for a detailed review of the possible and typical combinations of the progressive with lexical verbs

focuses on "the internal structure of a situation" in opposition to the perfective, which views the situation in its entirety [i.e. from the outside]" (Comrie 1976: 16).

In English the opposition between progressive and non-progressive aspect is a special case of the imperfectivity – perfectivity opposition which exists alongside other oppositions such as continuousness and habituality (cf. Comrie 1976: 3; 34):
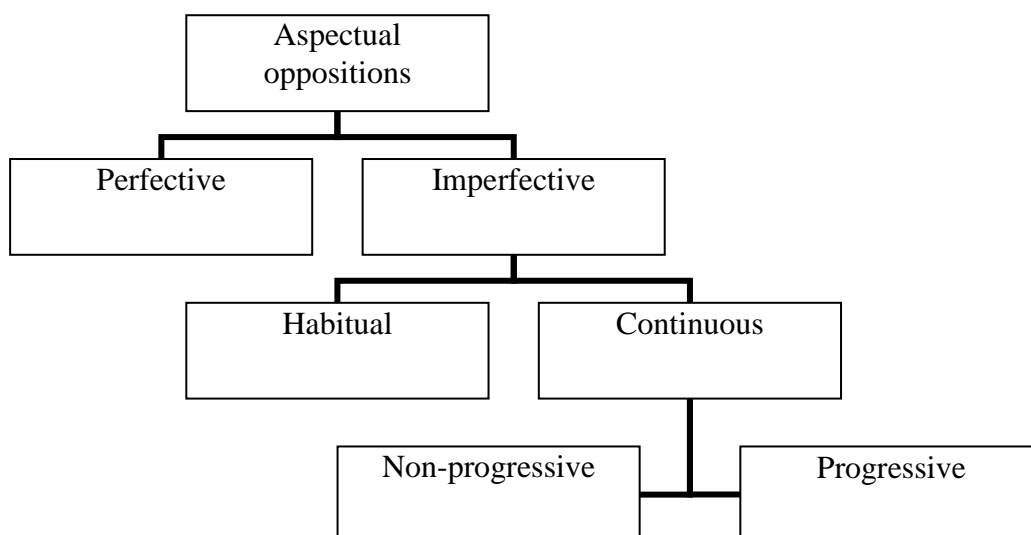
Figure 2.2: Classification of aspectual oppositions (adapted from Comrie 1976: 25)

Following Comrie's classification, continuousness is opposed to habituality which is characteristic for a longer period of time (e.g. 'John used to write', cf. Comrie 1976: 33), whereas progressiveness is a special kind of continuousness which is combined with non-stativity – the non-progressive being combined with stative verbs in turn (cf. Comrie 1976: 35). Bybee et al. (1994: 138) criticise Comrie for failing to delineate the continuous from the progressive precisely and note that the progressive is not necessarily restricted to non-stative verbs (cf. Bybee et al. 1994: 138 – 139). Comrie himself (1976: 38) admits that the meaning of the English progressive goes well beyond his general definition of progressivity as "the combination of continuous meaning and nonstativity" – and claims that it consists of a "general basic meaning which includes both progressive meaning and the various other meanings that the English progressive has" (ibid.: 38). This general basic meaning can express progressivity proper, a "temporary (contingent) state", or a "contingent habitual situation" (ibid.: 38) and can thus account for uses of the progressive involving stative verbs such as e.g. *I'm **understanding** more about quantum mechanics as each day goes by* (ibid.:

36, my emphasis)[12]. Comrie (1976) enumerates and exemplifies further non-progressive, non-aspectual uses of the progressive (e.g. "She's always buying far more vegetables than they could possibly eat" or "I've only had six whiskies and already I'm seeing pink elephants", ibid.: 37) and observes that the meaning of the English progressive has developed from purely aspectual to "a more extended meaning range" (ibid.: 39) which includes the combination with lexical verbs traditionally seen as "anomalous" in the progressive (e.g. *know*) (ibid.: 39). These anomalous combinations of the progressive with non-progressive verbs, the different meanings and meaning nuances of the progressive in modern English, as well as the semantic changes in its use over the past century will be discussed in greater detail in section 2.2.1. The next section will deal with the debate on the status of the English perfect as an aspect, tense or a third type of grammatical category.

2.1.4. The Perfect – Aspect, Tense or neither?

The status of the English perfect as a verbal category and in particular the semantics of the present perfect have long been the apple of discord for researchers. Comrie argues that "[t]raditionally, in works that make a distinction between tense and aspect, the perfect has usually, but not always, been considered an aspect, although it is doubtful whether the definition of aspect […] can be interpreted to include the perfect as an aspect" (Comrie 1976: 6). The reason behind the ongoing controversy on the nature of the perfect as an aspect, tense or a third type of category lies in the fact that the perfect says nothing about the internal temporal constituency of the situation, but rather "indicates the continuing present relevance of a past situation" (Comrie 1976: 52). The English perfect is formally realised by the auxiliary verb *have* followed by the past *–ed* participle, as in e.g. *I have walked the extra mile* (e.g. Biber et al. 1999: 460); in most general terms, it expresses a relationship between two time points – one describing the state resulting from a prior situation, and the time of that prior situation (cf. Comrie 1976: 52). Therefore, the perfect has to be held apart from the perfective aspect, which only "signals that the situation is viewed as bounded temporally" (as opposed to the imperfective aspect), but does not specify a temporal relationship between two points or intervals (Bybee et al. 1994: 54).

From an aspectual point of view, the perfect views a situation "from a particular perspective, namely from the perspective of the time when a result yielded by, or the

---

[12] a detailed review of the different meanings of the English progressive, the verbs it typically combines with and the lexical and other restrictions on its use will be delivered in section 2.2.1.

relevance of, an anterior situation expressed by the perfect form is perceptible" (Declerck 2006: 37). Declerck notes this perspective still refers to a temporal viewpoint, "i.e. a 'time of orientation' to which the situation expressed by the verb phrase […] is anterior" (ibid.: 38) and argues that the expression of anteriority with respect to an orientation time is a question of tense and not of aspect, since the choice between a perfect and a non-perfect form is not the same as the choice between a progressive and a simple form (cf. ibid.: 38). Binnick (1991: 264) observes that four different semantic theories of the perfect have emerged out of this temporal-aspectual controversy:

(1) the perfect "as an indefinite past as opposed to the definite preterite"
(2) the perfect as a current relevance past
(3) the perfect as an "extended now" past
(4) the perfect as a "past tense embedded within the scope of another tense, a kind of relative tense" (Binnick 1991: 264).

The fourth definition – the perfect as a relative temporal category (in contrast to the absolute tenses such as the simple past, cf. Declerck 2006: 212) has found support by a number of researchers like Bybee et al. (1994), who claim that "[a]nteriors (or "perfects", as they are often called) … are relational: an anterior signals that the situation occurs prior to reference time and is relevant to the situation at reference time" (Bybee et al. 1994: 54). Consequently, Bybee et al. (1994) and Kortmann (1995) maintain that the perfect can neither be classified as a tense nor as an aspect, but rather as a verbal category of its own (i.e. called anterior), "which marks the existence of an anteriority relation between a situation and a reference time" (Kortmann 1995: 186).

From a semantic point of view, both Kortmann (1995) and Klein (1992) call for a compositional analysis of the perfect (and the present perfect in particular), which incorporates several components and is loosely based on Reichenbach's three-parameter temporal model (see section 2.1.2.), involving the speech time S, the event time E and the "hotly debated reference point" R (Kortmann 1995: 185). On the basis of these three parameters Kortmann (1995: 185) defines the perfect as an order relation between E and R, such that E precedes R; whereas Klein (1995: 144) characterises the perfect is a relationship between the time of the assertion T-ASS following the time of the situation T-SIT (T-ASS AFTER T-SIT). As argued in section 2.1.2., both Klein's and Kortmann's terminologies refer to abstract semantic relationships which need to be kept apart from inflectional forms (cf. Klein 1995: 144). Similarly, Comrie (1976: 53) notes that not every form labelled "perfect" expresses perfect meaning – notably, the perfect in many languages (including German, see

section 2.3.) has extended its meaning to cover a variety of non-perfect uses like e.g. traditional definite past-time uses. Since the present corpus-based study adopts a form-oriented approach, a semantic analysis of the perfect in learner and native writing can only be carried out once the perfect forms have been identified in EFL learners' writing: non-perfect uses of these perfect forms, as well as perfect uses of non-perfect forms can only be analysed afterwards, on the basis of an initial formal analysis.

To summarise, there are different approaches to aspect and aspectuality, and especially with regard to the various formal realisations of aspectuality in English, as well as their exact number and meanings; all in all, Tobin (1993: 3 – 4) concludes that English aspect is a fuzzy and complex phenomenon which often defies description:

> English is notorious […] for expressing aspectuality in very many diverse ways which break the barriers between rigid traditional categories of tense and aspect, lexicon and grammar, syntax and semantics, and aspect and Aktionsart, thus making 'aspect in English' a particularly challenging area of research as well as fertile ground for comparing and contrasting alternative linguistic theories.

The following sections (section 2.2.1. and 2.2.2.) will drift away from the theoretical and semantic discussion of the role of aspect in English and will review some of the most recent form-oriented empirical studies, grammars and reference works exploring English aspect forms, their meanings and distribution across varieties and registers of spoken and written English. The final part of this chapter (sections 2.3. and 2.4.) will offer a brief contrastive analysis between the different realisations of aspect in German and Bulgarian as native languages of the EFL learners in the present sample and English as their target language, and will suggest possible difficulties that Bulgarian and German EFL learners may experience when using English aspect.

2.2. Aspect in Recent Empirical Grammars of English

The present section will review the treatment of grammatical aspect in some of the most recent corpus-based English grammars – specifically focussing on the widely-quoted standard reference works of the "Quirk fleet" (Görlach 2000: 260 in Mukherjee 2006: 337) – the *Comprehensive Grammar of the English Language* (CGEL) (Quirk et al. 1985) and two further usage- and corpus-based grammars which have been largely inspired by the CGEL and which "take […] the options offered by CGEL as [their] starting point for a quantitative

analysis" (Mukherjee 2006: 340): the *Longman Grammar of Spoken and Written English* (LGSWE) (Biber et al. 1999) and Mindt's *Empirical Grammar of the English Verb System* (EGEVS) (cf. Mindt 2000). What these three recently published reference grammars have in common is the fact that they rely on "empirical data from corpus-based analysis" (Conrad 2007: 55), using "computer-assisted techniques to analyze large, principled databases of naturally occurring language" (Conrad 2007: 55), although to a different extent. While the CGEL is the most intuitive and interpretative of the three, offering a "common core" (Mukherjee 2006: 338) evidence on the general use of grammatical aspect in English; and at the same time the first grammar of its kind to occasionally refer to the results of corpus-based studies (cf. Conrad 2007: 55; Mukherjee 2006: 33), the LGSWE and EGEVS focus on a number of under-researched quantitative perspectives on the distribution of aspect forms in English, identifying differences between different varieties of English (e.g. British English vs. American English) and different registers of English (e.g. spoken vs. written, fiction vs. non-fiction etc.). Mindt's grammar even further quantifies the distribution of the frequencies of particular meanings and functions of a specific aspect form (cf. Mindt 2000).

The data on grammatical aspect offered by these three corpus-based grammars are thus largely complementary – therefore, the researcher should make use of combined evidence on the basis of all three in order to ensure a comprehensive coverage of both the distribution and the patterning of meanings of aspect forms in English. Mukherjee (2006: 349) confirms that the LGSWE is "heavily dependent on the model and description set out in CGEL" and recommends a combined use of both grammars that should provide for a:

(1) comprehensive – and thus not necessarily and entirely corpus-based description of the grammatical structures that are possible and the demarcation from those structures that are not admissible in English
(2) the corpus-guided focus on routines (e.g. lexicogrammatical co-selections) and genre-specific trends that are typical of language use. (Mukherjee 2006: 349)

All three grammars agree on two aspect forms in Modern English: a progressive aspect, which in general terms views the action as incomplete; and a perfect aspect, which signifies past time with some current relevance (cf. Quirk et al. 1985: 189-190). In terms of their overall frequency distribution across different registers of Modern English, both the progressive and the perfect are rather infrequent phenomena and amount to less than 10% of all verb phrases in speech and writing (cf. Biber et al. 1999: 461), the majority of the finite verb phrases being marked for the simple aspect which is "overwhelmingly the preferred option" (Biber et al. 2006: 63).
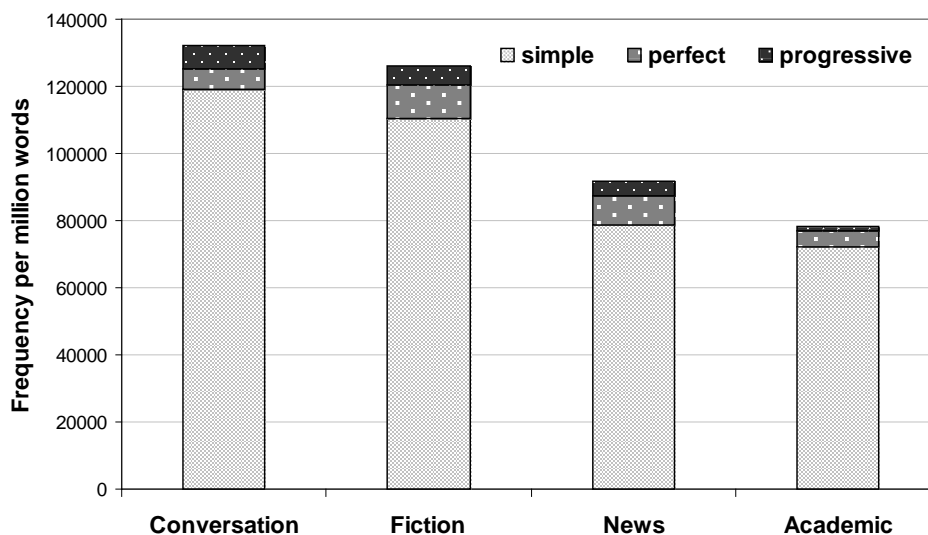
Figure 2.3. Distribution of the simple, perfect and progressive aspects across four registers of English (adapted from Biber et al. 1999: 461)

The following two sections will deal with the meanings and frequency distribution of the progressive and the perfect individually.

2.2.1. The Progressive in Recent Studies of English: Meanings and Frequency Distribution

The ongoing controversy about the nature of the perfect as an aspect, tense or a third type of category has not affected the progressive – like most theoretical accounts of the progressive, the majority of the empirical studies and usage-based grammars of English define the English progressive as an aspect and investigate it in a range of diachronic and synchronic frameworks of analysis in terms of its meaning variation, frequency of use and discourse functions (e.g. Comrie 1976, Quirk et al. 1985; Biber et al. 1999; Mindt 2000 etc.). This section will review the major functions and frequency distribution of the progressive aspect across different registers and varieties of present-day English.

From a semantic point of view, the central meaning of the progressive aspect on which all recent studies agree is "a happening in progress at a given time" (Quirk et al. 1985: 197). Quirk et al. (1985: 198) identify three components which constitute the meaning of the progressive:

(1) the happening has duration
(2) the happening has limited duration
(3) the happening is not necessarily complete

Thus, a sentence like 'I was reading a novel yesterday evening' (Quirk et al. 1985: 198) in the past progressive implies an action which had a limited duration and which is not necessarily complete – the person has not necessarily finished reading the novel. Furthermore, Quirk et al. specify three different kinds of progressives: state, event and habitual progressives (Quirk et al. 1985: 198 – 199). To illustrate, a sentence like 'We are living in the country' is a state progressive implying a certain temporariness of the situation, whereas 'Whenever I see her, she's working in the garden' is a habitual progressive implying a habit that repeats itself over a limited period of time. The sentence 'The referee is blowing his whistle' is an event progressive signalling an event that has not yet come to an end (all examples are taken from Quirk et al. 1985: 198 – 199). Quirk et al. identify three additional meanings of the progressive beyond temporariness – the use of the progressive to refer to the future, e.g. 'They were getting married the following spring', the use of the progressive as a marker of tentative wish, e.g. 'I was wondering if you could help me' and the use of the progressive with the auxiliary 'will' to imply that an action is taking place as a "matter-of-course" in the future, e.g. 'I'll be seeing you next week' (cf. Quirk et al. 1985: 210)[13].

Biber et al. (1999) in turn define the progressive aspect as a category of the verb phrase which "designates an event or state of affairs in progress, or continuing, at the time indicated by the rest of the verb phrase" (Biber et al., 1999: 460). Thus, the present progressive indicates an action or event currently in progress, and the past progressive an action or event that was in progress or about to happen (cf. Biber et al 1999: 470). In addition to the basic meanings outlined above, Mindt's (2000) more fine-grained division of the meanings of the progressive aspect includes as many as nine different meanings – 1) incompletion, 2) temporariness, 3) iteration/habit, 4) highlighting/prominence, 5) emotion, 6) politeness/downtoning, 7) prediction, 8) volition/intention and 9) matter-of-course, all of which are exemplified in table 2.6. (cf. Mindt 2000: 248).

---

[13] The meaning of future progressives goes beyond mere aspectuality (i.e. is part of modality); thus, future progressives will not be an object of investigation in the present study which focuses on aspect primarily

Meaning	Example
incompletion	It **was growing** cool, even cold, with the departure of the sun
temporariness	Carpenters **were putting up** wooden screens yesterday
iteration/habit	He **is** constantly **coming up** with bright ideas
highlighting/prominence	I've a feeling he's alive, and **I'm not thinking** of marrying anyone else
emotion	The sport **is hoping** to gain Olympic status
politeness/downtoning	I **am wondering** if you have any idea what it could be?
prediction	The subsidy **is being** withdrawn next year
volition/intention	**I'm going** to Paris for the weekend
matter-of-course	**He's writing**, of course, from the standpoint of his faith

Table 2.6. Meanings of the progressive in isolation (adapted from Mindt 2000: 256 – 261)

Many of these meanings of the progressive can be combined in a single proposition – overlaps of two or more meanings in a single verb phrase are thus fairly common and can make it difficult at times to distinguish between the individual meanings (cf. Mindt 2000: 256). These corpus-based findings are in line with Comrie's observation that "it may well be that English is developing from a restricted use of the progressive, always with progressive meaning, to this more extended meaning range" (Comrie 1976: 39). Mindt emphasises that not all nine meanings are equally distributed – the most frequent meanings are incompletion (60% of all uses), followed by temporariness (36% of all uses) and iteration or habit (12% of all uses), all three of which often occur in combination with each other, mostly featuring incompletion and another meaning component (cf. Mindt 200: 256 - 257). The frequency distribution of meaning combinations featuring incompletion as a set meaning component and further meanings as variable components is presented in figure 2.4:
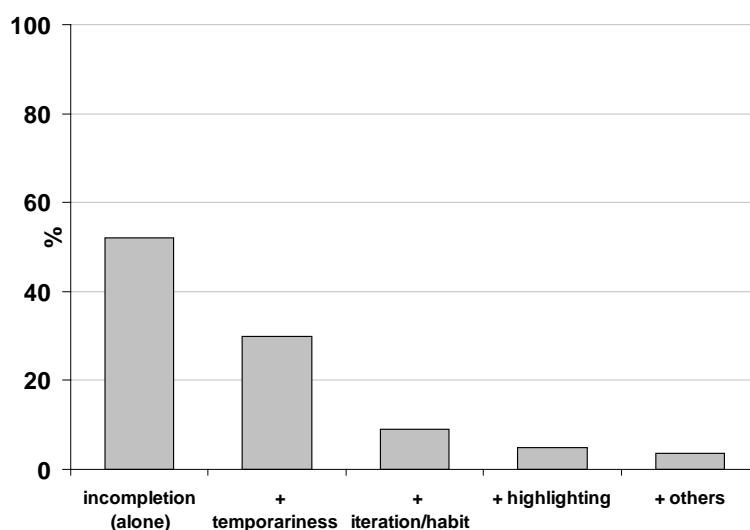
Figure 2.4. Distribution of incompletion and its combinations (adapted from Mindt 2000: 257)

The following two examples (taken from Mindt 2000: 257) illustrate the most frequent meaning combinations – incompletion with temporariness, and incompletion with iteration or habit:

2.1. Lucia **was thinking** how beautiful her mistress looked, and how cunningly the olive-green dress fitted her perfect figure. (incompletion + temporariness) (Mindt 2000: 257, original emphasis)
2.2. He saw the forester had closed his eyes and **was breathing** deeply. (incompletion + iteration/habit) (Mindt 2000: 257, original emphasis)

Further meaning combinations are not as frequent as incompletion with temporariness or incompletion with iteration/habit (see figure 2.4), but they also occur – examples 2.3. and 2.4. show a combination of incompletion with highlighting and incompletion with volition or intention (cf. Mindt 2000: 257):

2.3. He doesn't care at all what happens to the trees, or why **it's happening**, Julian said. (incompletion + highlighting)

2.4. I know what you're at, you'**re turning** it into a vegetable house. (incompletion + volition/intention) (Mindt 2000: 257, original emphasis)

In terms of the semantic association between lexical verbs and the progressive aspect, Quirk et al. (1985), Biber et al. (1999) and Mindt (2000) specify that the progressive occurs with a particular set of verbs – "verbs which mainly denote events (c. 87%)" (Mindt 2000: 264). Leech (1971: 19) argues that "most difficulties over the use of the [p]rogressive [a]spect arise with classes of verbs which are normally incompatible with the progressive". Quirk et al. (1985: 200 – 201) add that the constraints on the use of the progressive aspect with lexical verbs are further influenced by the verb complements (see also section 2.1.1); thus they differentiate between stative and dynamic situation types in general and a number of different subtypes of the stative and dynamic situations in turn, as illustrated in figure 2.5:

Stative
- A: Quality - e.g. *be tall, have two legs*
- B: State – e.g. *be hungry, be ill, love someone*
- C: Stance – e.g. *live, stand, lie, sit*

Dynamic
- Durative
  - Nonconclusive and durative
    - D: Nonagentive (goings-on) – e.g. *rain, snow, shine*
    - E: Agentive (activities) – e.g. *talk, hunt*
  - Conclusive and durative
    - F: Nonagentive (processes) – e.g. *ripen, grow up, improve*
    - G: Agentive (accomplishments) *write sth, eat sth, drink sth*
- Punctual
  - Nonconclusive and punctual
    - H: Nonagentive (momentary events) – e.g. *sneeze, explode*
    - I: Agentive (momentary acts) *tap someone, nod, fire a gun*
  - Conclusive and punctual
    - J: Nonagentive (transitional events) – e.g. *drop, receive*
    - K: Agentive (transitional acts) – e.g. *sit down, shoot, begin, stop*

Figure 2.5. Situation types (adapted from Quirk et al. 1985: 201)

Within the stative category, they claim that the progressive is not acceptable with the majority of the stative situation subtypes (e.g. 'We are *owning a house in the country', ibid.: 1985) and draw a distinction between qualities such as 'Mary is Canadian' and states such as 'Mary has a bad cold'. States can be further divided into "private states" or intellectual states (e.g. *know, believe*), states of emotion or attitude (e.g. *wish, want, like*) and states of perception or bodily sensation (e.g. *see, hear, smell, hurt, itch*) (all examples are taken from Quirk et al. 1985: 198 – 202). The third category within the stative situation types is "stance" and includes verbs like *live*, *stand* or *lie*, which can be used with the progressive to express a temporary state (e.g. 'James is living in Copenhagen'), and with the non-progressive to express a permanent state ('James lives in Copenhagen') (cf. Quirk et al. 1985: 206). In contrast to the restrictions on the use of the progressive with stative situation types, all dynamic situation types can be combined with the progressive according to Quirk et al., "but they have various implications for the interpretation of the progressive" (Quirk et al 1985: 207). Thus, durative situation types can be divided into goings-on (e.g. 'The engine was running smoothly'),

activities (e.g. 'The children are playing chess'), processes (e.g. ('The weather is getting warmer') and accomplishments (e.g. 'Jill is knitting herself a sweater') (all examples are taken from Quirk et al. 1985: 206 – 208). The last category includes punctual situation types such as momentary events (e.g. *nod*, *jump*) and transitional events (e.g. *arrive*, *die*). These can occur in the progressive, but they either imply a certain repetition of the action or event (e.g. "John is nodding his head"), or a "period leading up to a change of state" (Quirk et al. 1985: 209, see also section 2.1.).

Along similar lines, Biber et al. (1999) argue that the progressive commonly occurs with dynamic verbs from several semantic domains: verbs referring to activities and physical events, verbs referring to communication acts, verbs referring to mental attitudinal states or activities, verbs referring to perceptual states or activities, and verbs referring to static physical situations.

Semantic domain	Progressive verbs (> 50% in the progressive)
Activities and physical events	bleed, chase, shop, starve, dance, drip, head (for), march, pound, rain, stream, sweat, bring, buy, carry, come, cry, do, drive, eat, give, go, laugh, leave, make, move, pay, play, run, take, walk, work
Communication acts	chat, joke, kid, moan, scream, talk, ask, say, speak, tell
Mental/attitudinal states or activities	look forward, study, hope, think, wonder
Perceptual states or activities	look, watch, feel, stare, listen
Static physical situations	lurk, wait, sit, stand, wear, hold, live, stay

Table 2.7. Common lexical verbs in the progressive (adapted from Biber et al. 1999: 471 – 472)

In addition, Biber et al. (1999) observe that contrary to previous accounts of the progressive stating that it can combine freely with dynamic verbs, not all dynamic verbs can occur in the progressive; likewise, some stative verbs occur predominantly in the progressive (cf. Biber et al. 1999: 472 – 473). Thus, stative verbs such as *hope* or *think* involve a human subject who is also an active agent of the action and can therefore occur in the progressive without sounding awkward, whereas verbs like *want* or *desire* are "expressing a state experienced by someone" (Biber et al. 1999: 473) and are typically not used with the progressive, e.g. '*I am wanting to help'. Likewise, stative verbs such as *stay*, *wait*, *sit* and *stand* often have a limited duration and can occur in the progressive aspect; on the other hand, dynamic verbs which refer to an action which takes place instantaneously and has no duration rarely occur in the progressive, e.g. 'The man threw me/was throwing me out of the bus' (cf.

Biber et al. 1999: 474). A summary of the least frequent verbs used with the progressive is presented in table 2.8.

Semantic domain	Non-progressive verbs (< 2% in the progressive)
Activities and physical events	attain, award, dissolve, find, frighten, invent, rule, shut, shrug, smash, suck, suspend, swallow, throw, trap
Communication acts	accuse, communicate, disclose, exclaim, label, reply, thank
Mental/attitudinal states or activities	agree, appreciate, associate, attribute, base, believe, conceive, concern, conclude, correlate, delight, desire, know, like, reckon, suspect, want
Perceptual states or activities	detect, hear, perceive, see
Facilitation/causation or obligation	convince, entitle, guarantee, incline, induce, inhibit, initiate, inspire, interest, mediate, oblige, promise, prompt, provoke, render

Table 2.8. Lexical verbs rarely used in the progressive (adapted from Biber et al. 1999: 471 – 472)

Similar to Quirk et al. (1985) and Biber et al. (1999), Mindt (2000) identifies six different lexical verbs as the most frequent verbs occurring with the progressive and covering thus c. 60% of all progressive cases: the verbs *go* (as a main verb), *do*, *get*, *come*, *try* and *look*, in addition to other frequently recurring verbs such as *make*, *work, take, talk, wait, think, sit, begin, stand, say* and *become* (cf. Mindt 2000: 264).

In addition to the lexical preferences of the progressive outlined above, the three recent grammars as well as numerous further corpus-based studies compare its frequency of occurrence and register distribution across different varieties and registers of English (Quirk et al. 1985; Biber et al. 1999; Mindt 2000; Smitterberg 2005; 2008; Mair and Hundt 1995; Mair 1997; Nesselhauf 2007; Leech and Smith 2006; Smith and Rayson 2007; Hundt 2009; van Rooy 2006; 2008 among others). Quirk et al. (1985) are among the first grammarians to work with large-scale empirical data who identify the progressive as an "infrequent phenomenon" (Quirk et al. 1985: 198), claiming that it occurs in less than 5% of all verb phrases in present-day English (cf. Quirk et al., 1985: 198); Biber et al. (1999: 461) and Mindt (2000: 248-249) also confirm this finding. Albeit relatively infrequent, the progressive has been increasing in spoken and written registers of English ever since the 19th century (cf. Smitterberg 2005; 2008). Smitterberg (2005) shows in a comprehensive study of the progressive in 19th century British English that there is a steady increase in the progressive, although not in all his types of progressives and not in all genres he investigates (cf. Smitterberg 2005: 243 – 248; Smitterberg 2008: 268 – 269). Similarly, Mair and Hundt (1995) analyse the development of the progressive in later 20th century British and American English and identify an increase in the use of the progressive in written registers of English

over a period of thirty years (between 1961 and 1991). They account for this increase in terms of the gradual "colloquialisation" of British and American news writing – a process whereby a certain linguistic feature typical of conversational speech rather than writing suddenly becomes more frequent in written registers (cf. Mair and Hundt 1995: 225 – 226). Leech and Smith (2006) also confirm an increase of the progressive in written English by c. 30% and remark that there is "a tendency for spoken language habits to infiltrate the written language: colloquialisation" (Leech and Smith 2006: 198), claiming that this trend is sometimes accompanied by "Americanization" or the influence of American English usage "leading the way" in grammatical change in progress (Leech and Smith 2006: 199). Smitterberg (2008) also explains the increase of frequency of the progressive in the 20th century British and American English with the "colloquialisation of genre norms, which in turn has been linked to the democratisation of discourse in post 1945 Western society" (Smitterberg 2008: 269). Smith and Rayson (2007) establish an increase in the passive progressive and in particular the present passive progressive in present-day English, which they attribute to an overall increase in the progressive, whereas Mair (1997) in an earlier study asserts that the progressive is increasing as part of a change in progress and suggests that it is also taking on 'new' uses with stative verbs (cf. Mair 1997: 197). Along similar lines, Nesselhauf (2007) argues that there is "probably also [an] increase in the range of possible verbs and an increase of the not-solely-aspectual progressive" (Nesselhauf 2007: 205). Mair and Hundt (1995) and Smitterberg (2008) remark that the progressive in present-day English functions increasingly as a stylistic device, bridging the gap between spoken and written language and following the trend towards orality in written language and in particular in the language of newspapers (cf. Mair and Hundt 1995: 225 – 226; Smitterberg 2008: 284).

All in all, the progressive can thus be said to have increased considerably in use ever since the 19th century, both in speech and in writing, and to have extended its "traditional" aspectual uses to include non-progressive uses and combinations with non-progressive verbs verbs. Quirk et al. (1985:202) acknowledge that "[s]ince the use of the progressive aspect has been undergoing grammatical extension over the past few hundred years, it is likely that its use is still changing at the present day, and that its description at any one time cannot be totally systematic".

In terms of its distribution across different registers in present-day English, the progressive is most frequently found in informal conversation and least frequently in academic writing or expository prose, occurring predominantly in the present tense and in

main clauses (cf. Biber et al. 1999: 461; Mindt 2000: 248; 265). Figure 2.6 shows the distribution of the present and past progressive forms in conversation, fiction, news and academic writing (cf. Biber et al. 1999: 462). While the present progressive is most common in conversation followed by news, the past progressive is most common in fiction followed by conversation. Mindt (2000: 248) presents a similar distribution of the progressive, although his normalised frequencies per thousand words differ from Biber et al.'s (1999) frequencies slightly, presumably because he observes the frequency of the progressive in only three different registers (spoken conversation, fiction and expository prose) in contrast to Biber et al.'s four registers – conversation, fiction, news and academic writing (cf. Biber et al. 1999: 462; Mindt 2000: 248).
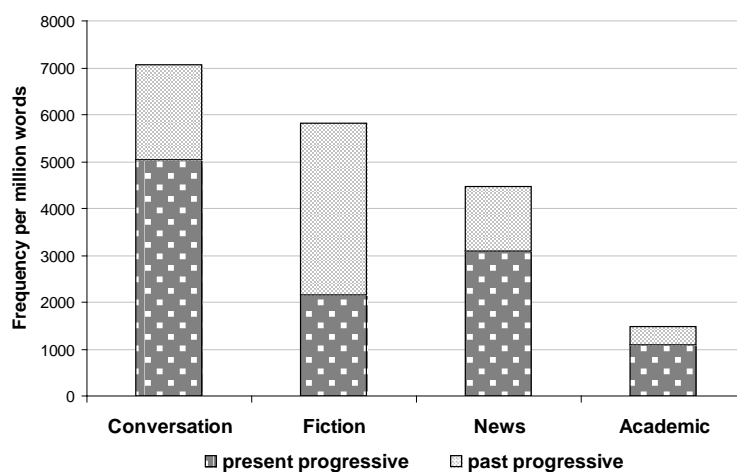
Figure 2.6. Frequency of the present and past progressives across registers (adapted from Biber et al. 1999: 462)

With regard to the distribution of the progressive across different varieties of English, Biber et al. (1999: 462) show that the progressive aspect is strongly favoured by American English in comparison to British English in the approximate ratio of 4:3 (cf. Biber et al. 1999: 461 – 462). This difference is especially significant in the case of British and American English conversation and to a lesser degree in the case of news (see figure 2.7).
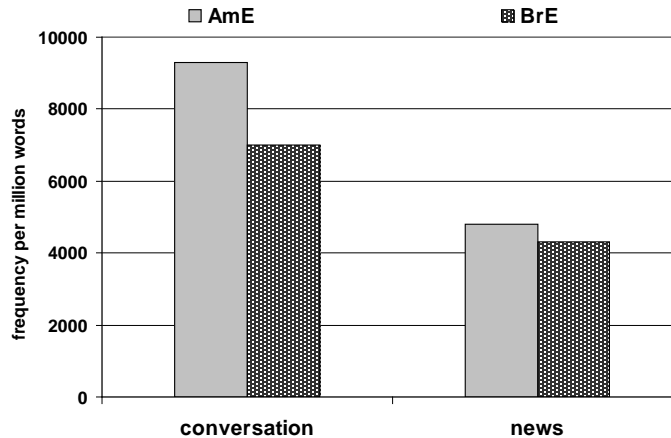
Figure 2.7. Frequencies of the progressive in BrE and AmE conversation and news (adapted from Biber et al. 1999: 462)

Apart from American English, other native varieties of English such as New Zealand English and especially second-language varieties of English (ESL) such as Indian English or Black South African English have shown a higher preference for the progressive in quantitative terms in comparison to British English (cf. Hundt and Vogel 2011: 155; Hundt 2009: 304 – 305; Gachelin 1997: 43 – 44). In addition to the already increased and further increasing use of the progressive, these varieties feature "very different ways of using the progressive construction that are not related to the core senses of the progressive aspect" (van Rooy 2006: 37) and that involve an extension of the progressive to stative verbs (cf. van Rooy 2006; 2008). The present study will not attempt to review all studies dealing with the extended use of the progressive in second-language varieties of English in Kachru's (1992) sense in detail, since such a review would go beyond the scope of the present investigation which focuses on learner varieties of English outside the Expanding Circle; still, similarities between learner frequencies and the patterns of use of the progressive commonly found in ESL varieties may occasionally be referred to in the empirical part of the study. The next section will survey the meaning variation and frequency distribution of the perfect aspect as outlined in recent corpus-based studies of English.

2.2.2. The Perfect in Recent Studies of English: Meanings and Frequency Distribution

The perfect aspect (and the present perfect in particular) is the second aspect form in English which has been identified as equally problematic for both EFL learners and grammarians (cf. Schlüter 2000). At the same time, the perfect aspect is the less empirically investigated form of the two – Bertus van Rooy (2009) argues that "the uses of perfect construction […] have

not attracted the same attention as the progressive" (van Rooy 2009: 310). The central meaning of the perfect aspect identified by the majority of the studies and summarised by Quirk et al. (1985) is the "anterior time" meaning, i.e. the "time preceding whatever time orientation is signalled by tense or by other elements of the sentence or its context" (Quirk et al. 1985: 190), within which the action described by the verb takes place. Similarly, Biber et al. (1999: 460) specify that the perfect aspect (which will be referred to as 'the perfect' for reasons of brevity) refers to "events or states taking place during a period leading up to the specified time" (Biber et al., 1999: 460). Thus, the present perfect has three basic meanings: "a state leading up to the present", "indefinite event(s) in a period leading up to the present", and "a habit (i.e., recurrent event) in a period leading up to the present" (Quirk et al., 1985: 192), all of which are derived from the anteriority meaning component. The following examples (taken from Quirk et al. 1985: 192) illustrate these three basic meanings:

2.5. That house **has been** empty for ages. (state leading up to the present)
2.6. Have you (ever) **been** to Florence? (indefinite event(s) in a period leading up to the present)
2.7. Mr Terry **has sung** in this choir ever since he was a boy. (habit i.e. recurrent event in a period leading up to the present) (Quirk et al. 1985: 192; original emphasis)

In addition, Quirk et al. (1985: 192) point out that the three meanings illustrated above correspond to meanings of the simple past, yet are different from them in several respects, mostly with regard to the perfect indicating "an *implicit time zone* which has not yet finished" (Quirk et al. 1985: 193, original emphasis) and the simple past describing a "situation that no longer exists or an event that took place at a particular time in the past" (Biber et al. 1999: 467). Hence, the action in a) corresponds to the 'state past' use of the simple past, but is different from it since the action described in the present perfect continues at least up to the present moment (in contrast to "The house was empty for ages – but now it's been sold" which signifies a completed action in the past); the action in b) corresponds to the 'event past' use of the simple past, but differs from it since it does not specify a definite point in time, while the action in c) corresponds to the 'habitual' past, but is different from it since it again describes an action which continues at least up to the present moment and has not yet come to an end (cf. Quirk et al. 1985: 192). Thus, the use of the present perfect has three major implications – 1) time zone leading up to the present, 2) recent event and 3) the result of the action obtaining at the present time (cf. Quirk et al. 1985: 193). Mindt (2000: 219) identifies five different meanings for the present perfect: 1) indefinite past (non-resultative), 2)

indefinite past (resultative), 3) recent past, 4) continuative past and 5) completion, all of which are exemplified in table 2.9:

Meaning	Example
indefinite past: resultative	he **has reached** the semi-finals
indefinite past: non-resultative	you have the most beautiful hair I **have** ever **seen**
continuative past	I **have looked** after my husband for seven years
recent past	they **have** recently **had** their third child
completion	teenage joyriders…often set fire to stolen cars when they **have finished** with them

Table 2.9. Meanings of the present perfect (adapted from Mindt 2000: 224)

These five meanings are not evenly distributed across all uses of the present perfect: the first three (the resultative and the non-resultative indefinite past and the continuative past), and in particular the resultative indefinite past account for the majority of the uses of the present perfect (over 90%, see figure 2.8) (cf. Mindt 2000: 224).
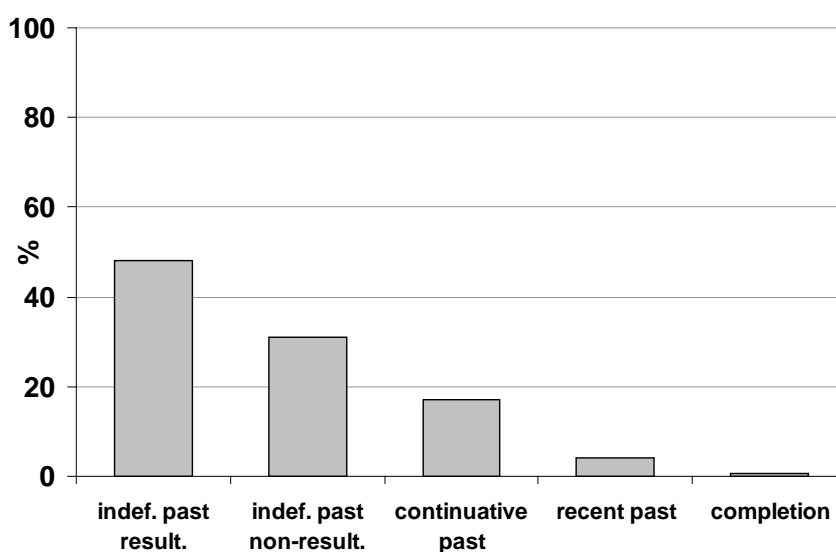
Figure 2.8. Distribution of perfect meanings (adapted from Mindt 2000: 224)

In terms of the lexical association between verbs and the present perfect, Biber et al. (1999) and Mindt (2000) list the most common verbs which frequently occur with the present perfect – predominantly event verbs (c. 73%) and state verbs (c. 15%) (cf. Mindt 2000: 227). The verb *be* as a main verb is the single most frequent verb in the present perfect in all registers apart from conversation, followed by *have* and *has/have got, do*, *go*, *see* and *come* (cf. Biber et al. 1999: 463; Mindt 2000: 227). Table 2.10 illustrates the most common verbs in the present perfect (with a frequency of over 40 times in a million words) occurring in at least one register (the verbs belonging to news reportage and academic prose occur in the present

perfect more than 25% of the time, cf. Biber et al. (1999: 463 – 464). In contrast, there are verbs which hardly ever occur in the present perfect (less than 2% of the time), such as e.g. *afford, aim, await, base, believe, compete, connect, depend, differ, matter, need* etc. (for a complete list of the verbs rarely occurring in the present perfect see Biber et al. 1999: 464).

Register	Present perfect verbs (> 40 times pmw)
news reportage	*agree, appoint, campaign, circulate, criticise, draft, experience, pledge, prompt, vow, witness*
academic prose	*criticise, document, implicate, master, report*
any register	*be, have, get, go, done, make, see, come, say, take, become, give, show, think, call, put, lose, win*

Table 2.10. Verbs that commonly occur with the present perfect (adapted from Biber et al. 1999: 463 – 464)

Alongside the present perfect, the past perfect has also been an object of investigation of recent empirical studies of English in terms of its meanings' distribution and lexical preferences. The past perfect is similar in meaning to the present perfect insofar as it can be regarded as an anterior version of the present perfect or a 'past-in-the-past' (cf. Quirk et al. 1985: 195), since it signals "a time before the past time referred to by the simple past tense" (Biber et al. 1999: 468). The three basic meanings typical of the present perfect (state, indefinite event and a habit) are also typical of the past perfect, as illustrated in examples 2.8 – 2.10 (taken from Quirk et al 1985: 195 – 196):

2.8. When we bought it, the house **had been** empty for several years. (state*)*
2.9. The goalkeeper **had injured** his leg, and couldn't play. (indefinite event)
2.10 It was foolish to fire McCabe: in two seasons, he **had scored** more goals than any other player. (habit) (Quirk et al. 1985: 196)

Mindt (2000: 237) identifies seven meanings of the past perfect, five of which coincide with his five meanings of the present perfect and two additional meanings – a definite pre-past for something which occurred at some definite point of time preceding the past, and a non-real past for something which might have occurred in the past, but did not occur. The additional meanings are illustrated in examples 2.11 and 2.12:

2.11. the incident **had happened** on the return journey (definite pre-past)
2.12. if he had gone to America they might have never met (non-real past, cf. Mindt 2000: 237 – 238)

The most common lexical verbs in the past perfect are very similar to those in the present perfect (*be, go, come, see*) – mostly verbs of physical activities (cf. Biber et al. 1999: 468;

Mindt 2000: 240) such as *leave, make, take, do, give, bring* etc. and mental perceptions such as *see, hear* and *know* (cf. Biber et al. 1999: 468).

Other lexical preferences for the perfect aspect concern the co-occurrence of the present and past perfects with adverbs of time and adverbial phrases – in the case of the present perfect, the difference to the simple past is often made explicit by the adverbs accompanying the main verb. Whereas verbs in the simple past often co-occur with a temporal adverbial phrase which specifies "a clear ending point before the present time" (Biber et al. 1999: 467), such as *in, during, yesterday, a few weeks ago, throughout*, etc., verbs in the present perfect are accompanied by adverbial phrases which do not signal an ending point or a definite point in time, but rather "the beginning point or the duration of the period of time" (Biber et al. 1999: 468), such as *already, since last January, now* etc. Mindt (2000) identifies the most common adverbs (accounting for over 45% of all adverbials) occurring with both the present and the past perfect as *already, never*, *just* and *always*, in addition to the adverbs which commonly occur with the present perfect such as *also* and *now*, or with the past perfect, such as *once* and *ever* (Mindt 2000: 229; 247). Nevertheless, although these adverbs have traditionally been employed as "trigger words" for the present perfect in EFL contexts, some of them like e.g. *already* and *always* have also been testified to co-occur equally frequently with other tense-aspect forms such as the simple present tense (cf. Voigt 2005: 128). In a more detailed corpus-based account of the adverbs and adverbial phrases co-occurring with the present perfect, Schlüter (2000; 2002; 2006) confirms Biber et al.'s (1999) and Mindt's (2000) results with regard to the most frequent adverbs co-occurring with the present perfect, claiming that they cover over 65% of all temporal modification (cf. Schlüter 2002: 311 – 312) and observing that single adverbs accompanying the present perfect are most common in informal registers of English, whereas prepositional phrases such as e.g. *for the moment, since the election* etc. are most frequent in more formal registers of English. The single most common adverbial phrase modifying the present perfect in all registers and both British and American English identified by Schlüter is the *(ever)since + temporal noun phrase* construction (cf. Schlüter 2002: 313). However, Schlüter (2000; 2002; 2006) remarks that temporal modification of the perfect is not as widespread as frequently suggested in EFL textbooks and classroom materials, quoting Peterson (1972: 3): "In teaching English as a foreign language we often teach the perfect constructions – especially the present perfect – in connection with certain adverbs and adverbial expressions. […] But in the material I analyzed the present perfect is used more often WITHOUT ANY ADVERBIAL EXPRESSION AT ALL." (Peterson 1970: 3 in Schlüter 2006: 141, original emphasis). On the basis of British

and American corpus data he proves that only about 33% of all present perfect verb phrases are temporally modified by an adverb or adverbial phrase, and that this proportion holds for both British and American English and for different registers of English (cf. Schlüter 2002: 313).

In terms of its frequency of use and register distribution in speech and writing, the perfect aspect is slightly more common than the progressive aspect, accounting for c. 5-10% of all verb phrases. Like the progressive, the perfect aspect occurs predominantly in the present tense in main clauses (cf. Mindt 2000: 229); unlike the progressive, there is disagreement as to whether the present perfect is more common in conversation or in specific kinds of writing (cf. Schlüter 2006: 139 – 140). Biber et al. (1999: 461) argue that the present perfect is most common in news (6.1 cases per 1,000 words, see figure 2.9), whereas other studies like Elsness (1997), Mindt (2000: 219) and Schlüter (2002: 109; 2006: 139) claim that the present perfect is most common in conversation (with varying frequencies of use ranging from 3.5 to 6 cases per thousand words, cf. Schlüter 2006: 140).
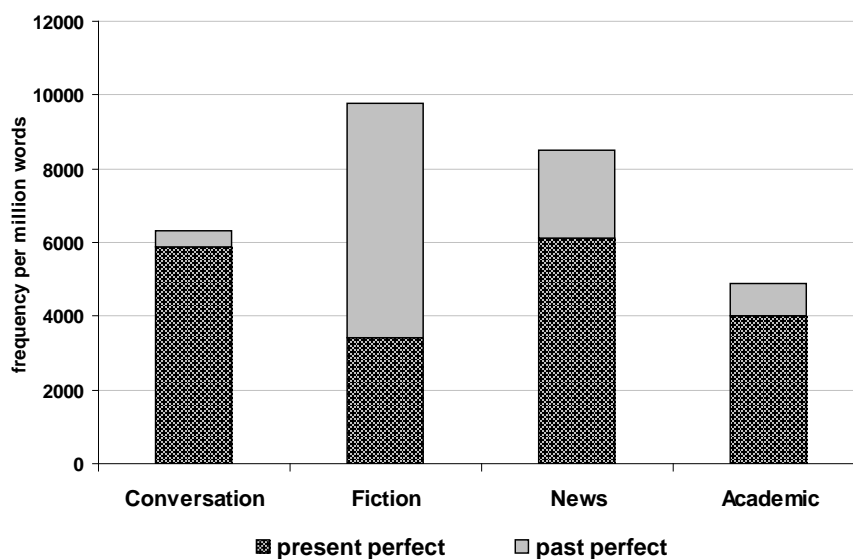
Figure 2.9. Frequency of the present and past perfect aspect across registers of English (adapted from Biber et al. 1999: 461)

Schlüter (2006) compares the results for the frequency of use of the present perfect in written and spoken registers of English obtained in several corpus-based studies (e.g. Elsness (1997), Biber et al. (1999), Mindt (2000) and Schlüter (2002)) and observes that although there are discrepancies between the normalised frequencies per thousand words presented by these four studies, all four studies are in agreement that the present perfect is most common in conversation, followed by expository prose and fiction (cf. Schlüter 2006: 147). The different

results obtained by Biber et al. (1999: 461) illustrated in figure 2.9. where 'news' leads 'conversation' with 1.1 occurrences per thousand words may be due to Biber et al.'s (1999) more fine-grained register division, which includes 'news' alongside 'expository prose', 'fiction' and 'conversation' in contrast to the studies quoted above, (e.g. Mindt 2000; Schlüter 2002) which consider only the last three registers[14]. Concerning the frequency distribution of the past perfect, the majority of the recent studies (e.g. Biber et al. 1999; Mindt 2000) agree that it is most common in fiction and least common in conversation. Further frequency details concern the distribution of present and past perfect verbs with elided auxiliary forms such as *'ve, 's* and *'d*, which are, as expected, most common in conversation, followed by fiction and expository prose (cf. Mindt 2000: 223; 241).

In terms of its variation across different varieties of English, the perfect shows the opposite tendency of the progressive insofar as it is more frequent in British English than in American English in the approximate ratio of 4:3 (cf. Biber et al. 1999: 461). As with the progressive, the differences between these two varieties are most marked in the case of 'news', although other registers also show a difference, but to a lesser extent (cf. Biber et al. 1999: 462 – 463). Hundt and Smith (2009) quote Strevens (1972) who claims that the American preference for the simple past over the British use of the present perfect as in e.g. 'Did you eat?' vs. 'Have you eaten?' has come to be "considered one of the shibboleths of transatlantic grammatical differences" (Hundt and Smith 2009: 45). Moreover, Mair (1997) suggests that simple past tense forms may "be encroaching on the past perfect and the present perfect" (Mair 1997: 197) not only in American English, but also in other varieties of English as part of an ongoing change in present-day English. Elsness (2008; 2009) argues that contrary to the "more general tendency [in European languages like French and German] of synthetic forms to be replaced by periphrastic constructions" (Elsness 2008: 229), the English present perfect is declining in due course of a linguistic change which is more advanced in American English than in British English (cf. Elsness 2009: 243 – 244). In a similar vein, Hundt and Smith (2009) identify a slight decrease in the use of the present perfect in both British newspaper writing and in American general prose over the past few decades of the 20th century (cf. Hundt and Smith 2009: 57). Marshall (1989) acknowledges that the so-called "colloquial preterite" or the "past tense [in American English] may be [used as] an informal

---

[14] on the other hand, Biber et al.'s (1999) frequencies of use must be interpreted with a pinch of salt – several studies point to the shortcomings of the LGSWE which bases its quantitative and qualitative analysis on different datasets, some of them fairly small with an unspecified design (cf. Mukherjee 2006: 345). Schneider (2001: 139 in Mukherjee 2006: 344) criticises the composition of Biber et al.'s samples which "is quite uneven" and which "conceal[s] a great deal of internal variation by topic, sociolinguistic background, etc.

alternative to the present perfect" (Marshall 1989: 307), even in textbooks and classroom materials within an EFL context – a proposal which mirrors the tendency of American English to use the same verb forms for both the preterite and the past participle (e.g. verbs like *burn*, *dream* and *leap*), as well as to regularise irregular verbs in colloquial usage (cf. Elsness 2009: 244). As with the progressive, a detailed survey of the diachronic development of the perfect aspect in English and its frequency and meaning variation across different Outer Circle varieties goes beyond the scope of the present study; therefore, the Expanding Circle will still remain the major object of research, although occasional references to the meaning and distribution patterns typical of the perfect in Outer Circle varieties of English may still be made in the following chapters. All in all, in terms of their development over the past few centuries, both English aspect forms have undergone and are still undergoing major changes in their frequencies and patterns of use; therefore, a contrastive analysis of learner use of the progressive and the perfect against a framework of such an ongoing change brings valuable insights not only into second-language research, but also into corpus-based research on temporality in general. The next two sections will present a brief contrastive analysis of aspect as a grammatical category in English and the aspectual systems of German and Bulgarian as native languages of the EFL learners in the present study.

2.3. Contrastive Remarks on Aspect: German vs. English

This section outlines the major similarities and differences between the aspectual systems of German as a native language of one of the learner groups in the present study and English as the target language of this learner group. Since the English aspectual system has been discussed in greater detail in the previous sections, this section will focus predominantly on the aspectual system of German in contrast to English and will not attempt to present a holistic contrastive analysis between the English and the German verb systems. Hahn (2007: 57) acknowledges that "the difference between tense and aspect […] has always been problematic for German learners [of English]", which she explains with the considerable differences between the English and the German aspectual systems. With regard to the German aspectual system, Löbner (2002: 373 – 375) states that "aspectual distinctions are not consistently marked in German" (Löbner 2002: 374) and that there is no distinction between imperfective and perfective aspect, but only between perfect and non-perfect aspect in Standard German. In contrast to English where the progressive aspect is fully grammaticalised, "[f]ully grammaticalised progressives are not particularly frequent in

Europe, with the exception of an 'Atlantic' area comprising the Iberian Peninsula, the British Isles and Iceland" (Dahl 2000: 21), and German makes no exception to this rule. Therefore, the progressive as a grammatical category is absent in German or "not grammaticalised to the same extent as in English" (Ebert 2000: 607). German grammars often subsume explanations of progressivity in German under headings such as "durative aspect" and *Aktionsart* (Ebert 2000: 605), presenting a variety of lexical means and expressions which are used to render "what is encoded by the progressive in English" (König and Gast 2009: 92). Filip (1989) argues that German linguistics has a long tradition of *Aktionsart* research which is "mainly understood as the lexicalisation of the relevant semantic distinctions by means of derivational morphology" (Filip 1989: 263) and claims that Aktionsart in German also operates by means of temporal adverbials (cf. Filip 1989: 263).

To illustrate, progressivity in German can be realised with the help of various temporal adverbials and prepositional phrases such as *gerade, nun, jetzt, zur Zeit, momentan, im Augenblick, im Moment, vorläufig, allmählich, zunehmend* etc., which are also recommended as translational equivalents for the English progressive (cf. Königs 1995). In addition, several periphrastic constructions realising progressivity which consist of the verb *to be* (*sein*) + a locative or a prepositional phrase, such as the preposition *am/bei* in combination with a nominalised verb (e.g. *Karl ist am/beim Arbeiten*), the adverb *dabei* + infinitive construction (e.g. *Ich war dabei, meinen Aufsatz abzuschließen, als du kamst*), or the preposition *im* + nominalised verb (*Diese Dinge sind im Kommen*) are among the commonly presented options of realising progressive meaning in German (cf. König and Gast 2009: 92 – 93). These "emergent" forms of the German progressive are especially common in varieties of German such as Rhine and Ruhr German (cf. Andersson 1989: 95 – 96) and are part of a larger trend of Germanic languages to realise progressivity with a number of periphrastic constructions that are on the verge of grammaticalisation (cf. van Pottelberge 2007: 112). In general, these periphrastic constructions are considered as forms of durative *Aktionsart*, rather than grammatical aspect (cf. Andersson 1989: 96) and serve as means for "explicit stressing of the actual going-on of a situation functioning in the text as background for a situation which carries the narrative on" (Andersson 1989: 105). König and Gast (2009) illustrate the frequency of the four different periphrastic constructions in German which render progressivity using the following scale (cf. König and Gast 2009: 93):

$$am + V_{nom} > dabei + \textbf{infinitive} > beim + V_{nom} > \textbf{im} + V_{nom}$$

These four periphrastic constructions are not freely interchangeable: whereas the most frequent one – the *am* + V_{nom} construction has the widest semantic range of usage of all four constructions, the *dabei* + infinitive and *beim* + V_{nom} constructions are much less frequent, as well as inappropriate with inanimate subjects, e.g. "*Die *Preise sind beim Steigen/*dabei, zu steigen*" (cf. Andersson 1989: 97). Andersson remarks that English has a similar periphrastic construction to the German *am* + V_{nom} construction – e.g. *he is **a-hunting*** (cf. Andersson 1989: 97, my emphasis). Even though the *am*+ V_{nom} is the most frequent one, it is regarded as regional or colloquial, while the *im/bei/dabei* periphrases are regarded as part of the Standard German lexico-grammatical repertoire. In terms of their lexical preferences, the four constructions can be combined with all but stative verbs (cf. Andersson 1989: 98; Ebert 2000: 605). However, even though the *am*+ V_{nom} construction may be fairly frequent in conversational German, it is not obligatory – the unmarked plain verb form can always be used as its alternative, sometimes even without a temporal adverbial, as well as an alternative to the other three constructions of König and Gast's scale. Thus, depending on the context, the sentence '*Er arbeitet'* can also serve to signify an ongoing action and can always be selected as a translational equivalent of "he is working" (cf. Hahn 2007: 58; König and Gast 2009: 93; van Pottelberge 2007: 112). Conversely, the linguistic contexts where one of the four periphrastic constructions can be used are limited and include the "core contexts" of the English progressive, such as current happenings or temporal frames serving as a background to the main story line (cf. König and Gast 2009: 93). Therefore, König and Gast (2009: 93) admit that "what we find in German is modest beginnings of grammaticalisation of an aspectual opposition with several competitors for the status of [p]rogressive aspect, which manifest clear contextual restrictions". Therefore, even if German learners of English may be familiar with ways to realise progressivity in spoken German, , their attention should still be drawn to the specificity of the English progressive and they should"learn to recognise and use the language-specific category of grammatical aspect in English, matching the correct progressive form with the appropriate progressive meaning (studies dealing with the acquisition of the English progressive by EFL learners will be reviewed in greater detail in the next chapter).

Unlike the absent category of a fully-grammaticalised progressive aspect in German, there is a grammaticalised category "perfect" in German which is marked by a perfect form (cf. Löbner 2002: 373). Formally, the German *Perfekt* is related to the English perfect, showing a "parallel formal make-up" of a past participle of a lexical verb and an auxiliary *have* (*haben*) or *be* (*sein*) (cf. König and Gast 2009: 86). *Be*-auxiliaries are much less

common than *have*-auxiliaries, especially in northern and central varieties of German, where *be*-auxiliaries are particularly rare (cf. Klein and Vater 1998: 220). Similar to the English tensed and modal perfects, the *have*-auxiliary in German can be marked for the present or the past tense, or can be non-finite (cf. Klein 2000: 358). In contrast to English, the German *Perfekt* is ambiguous in its meaning: it can function both as a non-past perfect and as a past non-perfect (cf. Löbner 2002: 388; Stechow 2002: 393). Contrastive studies and grammars agree on the fact that the biggest difference between the English perfect and the German *Perfekt* lies in the use of the German *Perfekt* as a narrative tense and its co-occurrence with temporal adverbials referring to definite moments in the past (cf. Comrie 1976; König and Gast 2009; Löbner 2002; Klein 2000; Klein and Vater 1998). This development of the perfect as a narrative tense in German is particularly typical of spoken German, as well as of regional varieties of German such as Southern German, where the perfect has almost fully supplanted the German simple past (*Präteritum*) in the majority of its uses – a linguistic phenomenon known as *Oberdeutscher Präteritumschwund* (Klein 2000: 359). Along similar lines, Bybee et al. (1994) argue that "[i]n modern German, the anterior has extended its use and is taking over the functions of the past tense" (Bybee et al. 1994: 85). Thus, the German sentence '*Ich habe den Brief gestern um 10 abgeschickt*' (literally: '*I have sent the letter at 10 yesterday*') is perfectly acceptable and frequently used in German (cf. Klein 2000: 359).

König and Gast (2009: 86) distinguish between two main uses of the German *Perfekt* – 1) a resultative use and 2) a narrative use. The resultative use of the German perfect is equivalent to the English perfect insofar as it signals an indefinite past with an obvious result, as illustrated in example 2.13 (cf. König and Gast 2009: 86):

2.13. Schau mal, es hat geschneit! (König and Gast 2009: 86)

One difference between the German resultative perfect and the English present perfect concerns the temporal specification of the German perfect – since the German present tense (*Präsens*) can also have future time reference, the German *Perfekt* may refer to the present, past or even future: thus the sentence '*Morgen Abend habe ich dieses Kapitel abgeschlossen*' (König and Gast 2009: 87) refers to a future resultative event employing the *Perfekt* (cf. König and Gast 2009; Klein 2000; Klein and Vater 1998). The second type of use of the *Perfekt* which is becoming more widespread and which is also dramatically different from the meaning of the English perfect is the narrative use:

2.14.   Gestern sind wir ins Kino gegangen. Wir haben uns den neuesten Film vom Wim Wenders angesehen. Anschließend haben wir bei einem Italiener gegessen. (König and Gast 2009: 86)

In example 2.14, the German *Perfekt* combines freely with temporal adverbials signalling definite past moments that have come to an end like *gestern* ('yesterday'). König and Gast (2009: 86 – 87) argue that this development of the German *Perfekt* as a narrative tense is part of a language change in progress and claim that this narrative use is formally and stylistically marked, since more formal written genres in German still prefer the use of the *Präteritum*. Therefore, within its narrative use, the perfect in German is interchangeable with the *Präteritum*; in contrast, the *Präteritum* is inadmissible within the resultative uses of the German perfect, as illustrated in example 2.15:

2.15.   Unser Hund ist weggelaufen. Wir müssen schnell etwas tun. (cf. *Unser Hund lief weg, König and Gast 2009: 87)

Klein (2000: 359) states that "whenever a present situation is somehow presented as a result of a past situation, the *Perfekt* but not the *Präteritum* is possible", and only in these situations does the German *Perfekt* correspond to the English present perfect. Conversely, uses of the English perfect such as an experiential perfect signalling indefinite events leading up to the present (e.g. "Have you ever been to Paris?"), as well as continuative uses (e.g. '"I have looked after my husband for seven years") arenot always necessarily rendered by the German *Perfekt*, but rather by the *Präteritum* or the *Präsens*, as illustrated in examples 2.16 and 2.17:

2.16.   Warst du (je) in Paris?[15]
2.17.   Ich warte hier (schon) drei Stunden lang. (Klein and Vater 1998: 229 – 230)

In the case of the continuative use of the perfect, German can employ both the *Perfekt* and the *Präsens*; however, "only the *Perfekt* implies that the situation does not extend beyond the moment of speech" (König and Gast 2009: 89). Nevertheless, continuative uses of the English perfect are typically translated with the German *Präsens*, so that the sentence 'I have lived here for many years' is rendered as 'Ich lebe hier seit vielen Jahren' (König and Gast 2009: 89). This particular contrast between German and English often leads to errors made by German learners of English who tend to produce sentences like "I learn English since ten years" (Erling 2002: 8), in an attempt to transfer the continuative use of the German *Präsens* back to English, where the present perfect would have been the appropriate form for this particular function.

---

[15] Bist du je/schon in Paris gewesen? is, however, becoming increasingly more common

The ambiguity of the German *Perfekt* (which serves both as a genuine perfect and as a narrative past tense), together with its formal similarity to the English perfect, lies at the heart of the problem for German learners of English. Hahn (2007: 57) notes that German EFL learners experience difficulties differentiating between the simple past and the present perfect in English, since the "morphologically marked forms "past" and "present perfect" do not make any difference to the temporally encoded meaning [in German]". Therefore, German learners of English should be aware not to 'fall back' on the tense-aspect principles specific to their mother tongue, which will be "useless in the target language" (Hahn 2007: 57). The brief contrastive analysis of the aspect systems of German and English outlined in this section shows that although both German and English belong to the same language family and display some historical and formal similarities, they encode the universal concept of time differently – by employing linguistic means which occupy different positions on the lexis-grammar continuum – a finding which has to be borne in mind whenever learners (German learners of English or English learners of German) acquire the aspectual system of the respective second language. The next section will briefly illustrate the aspectual system of Bulgarian in contrast to English.

2.4. Contrastive Remarks on Aspect: Bulgarian vs. English

In contrast to the formal and to a certain degree functional parallels between German and English as members of the Germanic branch of Indo-European, Bulgarian as a Slavic language bears little formal and functional similarities to English. In terms of its verb system, Bulgarian shares many of the characteristics typical of the Slavic family of languages with other Slavic languages such as Russian (cf. Damova 1999: 143); most importantly, it has the Slavic type of aspect (cf. Dahl 2000: 21). This type of aspect, also commonly called "vid" (literally 'type', cf. Damova 1999: 148) in the literature is "realised as an affixal and morphosyntactic category" (Lindtstedt 1985: 39) "by clear-cut morphological distinctions such as prefixes or different verb forms" (Tobin 1993: 3). In general, "vid" refers to a classification of each verb stem as either imperfective or perfective – both alternative verb forms constitute a "systematic grammatical opposition" of each verb form (Tobin 1993: 3). The perfective form of the verb is the marked member of the opposition and specifies that the action is complete, whereas the imperfective verb form is the unmarked form (cf. Scatton 1984; Scatton 2002; Lindstedt 1985). These two verb stems form an aspectual pair and relate

"the same verbal action but from a different perspective" (Scatton 1984: 318), as illustrated in the following examples from Bulgarian:

2.18.   Вчера *купих* много книги..)
yesterday buy PST 1SG many book PL
'Yesterday I bought many books'
2.18.1. Вчера цял ден *купувах* книги. yesterday whole day buy IMPST 1SG many book PL
2.19.   Yesterday the whole day I was buying books. (cf. Scatton 1984: 318, my translation)

Perfective st*ems are* usually formed from no*n-prefi*xed imperfective stems by means of prefixation (e.g. *piša* (imperf.) – **na**piša (perf.), 'write/write down' or suffixation (e.g. *padam* (imperf.) – *pad**n**a* (perf.), 'fall') (cf. Scatton 2002: 212, my emphasis). In addition, the so-called 'secondary imperfectives' (cf. Scatton 2002: 212; Lindstedt 1985: 41) can be formed from almost every perfective verb stem with the help of imperfectivising suffixes: thus the suffix – *(a)va* is the most productive suffix in Bulgarian which can be used to produce secondary imperfective verbs, e.g. *izbroja* ('count', perf.) → *izbroj**ava**m* (imperf.) (cf. Scatton 2002: 213). The perfective/imperfective division of verb stems is typical of all Slavic languages and has also been the object of long debates with regard to its exact nature as an aspect, *Aktionsart* or another type of category (cf. Binnick 1991: 148); still, the majority of the contrastive studies on aspect in Bulgarian (and especially the more recent ones) agree on the fact that this opposition is part of derivational morphology and thus "deeply rooted in the lexicon and […] fundamentally ascribable […] to the category of actionality, rather than aspect proper" (Bertinetto and Delfitto 2000: 190). Therefore, the Bulgarian perfectivity/imperfectivity distinction of verb stems can be considered either as a situation type of aspect (*Aktionsart*) or as a third type of category, which is distinctly different from the English periphrastically-realised aspect[16].

In addition to the generic Slavic-type classification of verb stems as either perfective or imperfective, Bulgarian has further means of realising imperfectivity with reference to the past – by the imperfect past tense, which is opposed to the perfective or the aorist past. Aronson (1984: 275) claims that "the aspect system of Bulgarian … [is] the richest of the Slavic languages because there are two distinct types of aspect oppositions: the opposition traditionally called perfective/imperfective and the opposition aorist/imperfect". The aorist and the imperfect forms of the verb 'to write' are illustrated in table 2.11.

---

[16] Binnick (1991: 148) notes that Slavic type of aspect should be kept apart from grammatical aspect and *Aktionsart*, and emphasises that "[v]irtually any Slavic verb may have either perfective or imperfective stems. It is an obligatory category of the Slavic verb and pervades the system of tense and aspect"

past aorist 'write'	past imperfect 'write'
писах / pisah	пишех / pišex

Table 2.11. Past aorist vs. past imperfect (adapted from Scatton 2002: 212)

Both the aorist past and the imperfect past forms are two synthetic forms which have been occasionally regarded as two separate tenses in Bulgarian (cf. Scatton 1984: 42 – 43; Lindstedt 1985: 65); nevertheless, Binnick (1991: 147) points out that this distinction has often been termed "aspectual" by scholars. The difference between the two forms lies in the fact that whereas the aorist past tense conveys "successive, independent major occurrences of the narration […] and denotes past actions completed at some definite, specific time" (Scatton 1984: 42 – 43), the imperfect past tense "relates past events which are in some sense subordinate to aorist events […] also used for repeated, habitual independent past events" (Scatton 1984: 43). Somewhat more complex are the different combinations between imperfective and perfective verbs with the aorist and the imperfect – all four combinations between the two past tense forms with imperfective and perfective verb stems such as perfective imperfects (e.g. *напишех/napišex* 'write up') and imperfective aorists (e.g. *написвах/napisvah*) are possible in Bulgarian – perfective imperfects are thus used to refer to subordinate actions of the main storyline, whereas imperfective aorists are used for "major past events the completion of which is not relevant for the narration" (Scatton 2002: 213). In both cases the tense (aorist or imperfect) overrules the aspectual morphological characteristic of the verb: perfectivity dominates in imperfective aorists, whereas imperfectivity dominates in perfective imperfects (cf. Comrie 1976: 32). Still, imperfective imperfects and perfective aorists are most common. Imperfective imperfects are also the only ones which occur in "single, independent verb phrases" (Scatton 1984: 323), such as example 2.20:

2.20. Той четеше/Toj četeše. ('He was reading') (Scatton 1984: 323)

In cases where the imperfect accompanies the aorist in a sentence and relates actions which run parallel to the main events narrated by the aorist, grammarians often translate the Bulgarian imperfect using the English past progressive, and the Bulgarian aorist using the English simple past (cf. Scatton 1984: 322 – 323). The similar use of the English past progressive and the Bulgarian imperfect past is the most straightforward functional similarity between the aspectual systems of the two languages with regard to the realisation of progressivity; yet it is an insufficient similarity, since the English progressive is a fully grammaticalised analytic form which expresses a specific type of imperfectivity (cf. Comrie

1976: 33), whereas the Bulgarian imperfective/perfective distinction, together with the aorist/imperfect tense opposition, constitutes a more complex aspectual system in which derivational and syntactic categories are intertwined to convey different subtypes of imperfectivity. Nevertheless, it is likely that Bulgarian learners of English may consider the English past progressive as an equivalent of the Bulgarian imperfect past tense; however, the consequences of equating the two forms are not clearly predictable and have not been addressed thus far in the literature. One possible consequence may concern Bulgarian EFL learners' intuitive reliance on a morphologically-coded imperfectivity to realise progressivity in English and their subsequent non-use of the English progressive in required progressive contexts.

In contrast to the non-existent category of a progressive 'proper', Bulgarian has a formally distinct perfect which is a compound analytic form. Lindstedt (2000: 371) remarks that Bulgarian belongs to the "maritime category" of languages in Modern Europe since "most of the languages and dialects with a stable perfect are situated on the fringe of the continent: the Baltic Finnic languages, Scandinavian languages, North German dialects, English, Portuguese, Spanish, South Italian dialects, Greek, Albanian, Macedonian and **Bulgarian**" (Lindstedt 2000: 371, my emphasis). The perfect in Bulgarian consists of the present tense forms of the verb "be" and the aorist participle of the main verb, e.g. *чел съм/čel sŭm* ("I have read"), where the verb "be" is inflected for person and number and the aorist participle is inflected for gender and number, e.g. (masc. *čel sŭm*/ fem. *čela sŭm*) (cf. Scatton 1984: 323). In terms of its function, the perfect in Bulgarian relates an action which has been completed in the past, but is in some ways relevant to the present moment (cf. Scatton 1984: 323; Scatton 2002: 211). The perfect is thus also known as an "indefinite past" (Andrejčin 1978 in Fici 2005: 36) and can also express a variety of further meanings, ranging from aspectual and temporal to attitudinal meanings (cf. Fici 2005; Fielder 1995). In terms of its aspectual meaning, the perfect with perfective verbs echoes the English "indefinite past with present result" perfect meaning, since it emphasises that the result of an action or an event is still relevant to the present. The perfect with imperfective verbs emphasises that "the subject has carried out such an action at some unspecified time in the past" (Scatton 1984: 323) and is thus similar to the non-resultative experiential uses of the English perfect. Examples 2.21 and 2.22 illustrate the resultative use of the Bulgarian perfect with a perfective verb and the non-resultative use of the Bulgarian perfect with an imperfective verb:

2.21. Купил съм си ново палто/Kupil sǔm si novo palto. ('[I] have bought a new coat') (cf. Scatton 1984: 324)

2.22. Чел ли си Пушкин?/Čel li si Pushkin? ('Have you read Pushkin?')

Hence, in contrast to German and other European perfects and similar to the English perfect, the Bulgarian perfect has retained some of its aspectual meanings like the experiential meaning exemplified above (example 2.22) – thus the English sentence "I have been to Paris" is commonly translated with the Bulgarian perfect equivalent (cf. Lindstedt 2000: 377). In contrast to German, where the perfect has developed into a more general narrative tense, the Bulgarian perfect is unusual in narratives (cf. Fici 2005: 38), although it may occur in "extended narrations of past events with present tense forms [...] in order to introduce prior actions which are relevant for the 'present' narrative moment" (Scatton 1984: 324). Furthermore, other uses of the Bulgarian perfect are related to a very different development, which has transformed the Bulgarian perfect into a category of evidentiality (cf. Fielder 1995; Lindstedt 2000; Fici 2005). Evidentiality (also called *preizkazni formi*, 'forms of hearsay', cf. Fici 2005: 39) refers to the meaning of the perfect which expresses that the speaker did not directly witness the action or event mentioned, and thus "indicates something about the source of the information in the propositions" (Bybee 1985: 184). On the formal side, evidential or indirect uses of the perfect (cf. Lindstedt 2000: 376) are difficult to hold apart from aspectual uses, since they are identical with them with the exception of the third person singular, where the auxiliary form of "be" in the evidential use is missing from the otherwise identical perfect construction – a phenomenon known as "0-auxiliary" or "auxiliary-drop" (e.g. '*Той мислил по този въпрос/Toj mislil po tozi vupros*' ('He 0-aux. considered (perf.) this question'), cf. Fici 2005; Lindstedt 2000). Due to the substantial formal overlap between perfect and indirect forms, researchers disagree whether to regard evidentiality as a fully-fledged category in Bulgarian; Lindstedt (2000: 376) argues that "[t]he distinction between the Perfect and the Indirective has been one of the most difficult questions for Bulgarian grammarians [...] and no definitive solution has been reached yet". In terms of meaning, evidential or indirect uses of the perfect signal not only that the speaker has not witnessed the event or action first-hand, but also the speaker's attitude towards the reported event or action – a characteristic of indirect forms which would fall under the subcategory of 'modality', rather than 'temporality' (cf. Fielder 1995: 585 – 586). In terms of their use in different genres of spoken and written Bulgarian, perfect forms are more typical of dialogues (cf. Fici 2005: 36), whereas indirect forms are more typical of narratives such as fairy tales (cf. Fici 2005: 39; Lindstedt 2000: 376 – 377), as well as scientific articles and reports whose authors quote the findings of other

scholars (cf. Fici 2005: 39). This latter use of indirect forms in Bulgarian may influence Bulgarian EFL learners' choice of tense-aspect forms when writing in English, inasmuch as Bulgarian learners may prefer to use the English perfect in argumentation at the expense of the simple present or the simple past, in an attempt to signal lack of "first-hand" evidence about certain facts or events. In sum, although Bulgarian as a native language is formally and genetically unrelated to English, there are some ever so slight functional parallels between the aspectual systems of the two languages, and in particular in terms of the use of the perfect aspect.

2.5. Summary

Since "aspect is perceived as one of the core areas of English grammar, and its mastery [in EFL] is regarded as *sine qua non* of the mastery of English" (Lorenz 2002: 132), a contrastive "mastery check" of the use of English aspect in learner language is certainly a new area of research worth delving into. Notably, there are no one-to-one parallels between English aspect and the lexical and grammatical means used to realise aspectuality in Bulgarian and German as native languages of the learner groups in the present study, a fact which poses additional difficulties for Bulgarian and German EFL learners. Put in a nutshell, the task for both Bulgarian and German EFL learners in using the English progressive and perfect aspect appropriately consists not only in learning to use the forms, but also in learning to map the appropriate functions onto the correct verb forms and to combine them with the correct tenses and adverbials in the appropriate contexts.

This task may pose different challenges for the two learner populations in the present study: while German is genetically close to English and a reliance on the formal similarities between the two languages may be productive in lexical terms (cf. Kellerman 1997: 288 – 289), falling back on the principles of use of the German *Perfekt* and using the equivalent present perfect form in English to narrate happenings from the past would be clearly counterproductive for German EFL learners. Moreover, German EFL learners' use of the English progressive as a new grammatical category is difficult to predict: German EFL learners could either fail to use the progressive in required progressive contexts, or they could use it more frequently, but less idiomatically than native speakers of English[17].

---

[17] Several studies comparing a variety of learner populations with Germanic native-language backgrounds show that the progressive tends to be used more frequently than required; for more detail on the overuse of the progressive, see chapter 4.

In sharp contrast to German, Bulgarian as a language which is genetically more distantly related to English poses no comparable difficulties in terms of misleading formal parallels between the two languages; however, functional similarities may still mislead Bulgarian EFL learners into using e.g. the English perfect as a modal/evidential category which marks indirectness in simple argumentation, rather than using it in its aspectual meaning to signal events from the recent past. Like with German EFL learners, the progressive is a new type of category for Bulgarian EFL learners too, since the corresponding imperfective verbs (expressing progressive meaning) in Bulgarian are part of the lexicon, rather than part of the grammar, and are realised through derivational affixes (cf. Scatton 1984: 318). As a result, instead of using the corresponding periphrastic English progressive, Bulgarian EFL learners may simply 'forget' to use progressive markers altogether, especially in present-tense contexts, where they may equate the *Aktionsart* of English verbs (e.g. atelic, dynamic verbs) with progressive uses, thus avoiding to inflect them for the progressive.

The brief contrastive analysis of the aspectual systems of Bulgarian and German as native languages and English as a target language presented in this chapter is barely sufficient to predict whether a closely-related native language such as German would necessarily mean better grammatical performance of German EFL learners in comparison to Bulgarian EFL learners at the same level of proficiency; nevertheless, it provides a suitable framework for comparison between the learners' performance in the empirical part of this study by delivering the diagnostic, rather than the prognostic tools. Lorenz remarks that irrespective of the mother-tongue background, "learners of English tend to see the [English] aspectual system as fundamentally 'English' […] and as perceptionally highly salient" (Lorenz 2002: 133), and are prone to notice and remember irregular patterns in authentic native-speaker use of the progressive and the perfect which may differ substantially from the norms of 'good English' they have been explicitly taught in school. Therefore, in addition to the mother-tongue backgrounds in the present sample, there are also a number of further learner- and learning-related factors like the learning environment, the teaching effects, the amount of target-language exposure etc., which need to be taken into account when comparing authentic aspect use by the two learner populations. These factors will be considered in greater detail in chapters 8 and 9. The next chapter will give an overview of the theoretical framework on the acquisition of aspect in English as a second language from the point of view of universal theories regardless of the native language, as well as from the point of view of the native-language influence on the use of temporality in L2 English.

3. Second Language Acquisition of Aspect in English: a Review

3.1. L2 Acquisition of Temporality

The acquisition of temporality in a second language and the emergence and development of aspect marking in particular have received considerable attention over the past thirty years in a variety of research contexts and theoretical frameworks defined as a "small, but active area of investigation in SLA" (Bardovi-Harlig 1994: 41). Apart from investigating verbal morphology in English as a second/foreign language, a substantial number of studies focus on a wider range of linguistic means used for the realisation of aspect in L2, ranging from lexical means like temporal adverbials to pragmatic means such as patterns of discourse organisation (Shirai 2009; Salaberry 2002a; Salaberry 2002b; Noyau 2002; Bardovi-Harlig 1999; Bardovi-Harlig 1994; Schumann 1987). Most studies investigating the second language acquisition of aspect in English have addressed three major research questions relating to the overall acquisition of tense-aspect morphology:

(1) How do learners acquire the English verb system? How do they acquire morphological distinctions between base forms and inflected forms?
(2) What stages of development are there in the acquisition of forms? Do some forms precede others? Which forms are acquired first and which last?
(3) How do learners acquire the function of these forms? How do they use them in the immediate linguistic context? Are there specific developmental stages in the acquisition of functions and if so, do they precede or follow stages in the acquisition of forms?

The last two questions refer to the twofold task of learners acquiring tense-aspect morphology: on the one hand, the acquisition of the correct verb forms, i.e. a "form-to-form mapping" (Housen 2002b: 155), and on the other, the acquisition of the functions and meaning of these forms, together with their temporal and discourse properties, and the subsequent appropriate use of these forms in specific linguistic contexts, i.e. the "form-to-function mapping" (Housen 2002b: 156).

The present chapter will review the major theoretical frameworks in the study of the second language acquisition of tense-aspect morphology, focussing on English as a target language and summarising the existing research methods, results and implications for learners

of English from a wide range of native languages, proficiency levels and acquisitional environments.

Studies exploring the development of markers of temporality in interlanguage fall into two major research strands: those focussing on form – the form-oriented approach; and those focussing on meaning – the concept-oriented approach (Shirai 2009; Bardovi-Harlig 1994; Bardovi-Harlig 1999; Bardovi-Harlig 2007). Studies following the form-oriented approach, which are also known as form-to-function studies, investigate how, when and where a specific linguistic form is acquired and used by second language learners. Studies following the concept-oriented approach use a broader research framework – they explore second language use from the perspective of a specific semantic concept, e.g. the concept of temporality, which they investigate by looking at its various realisations (lexical, grammatical, pragmatic etc.) and the order of their emergence in second language use. In doing so, the concept-oriented studies mostly rely on a qualitative analysis of the different realisations of temporal and aspectual relationships, including tense-aspect morphology. In contrast, the form-oriented studies rely predominantly on a quantitative analysis of the distribution of tense-aspect forms, which may (or may not) involve a further qualitative investigation of the use and meaning of these forms.

Since the present study is a corpus-based study on aspect use in learner writing, it will adopt primarily a form-oriented approach to aspect morphology as used by advanced German and Bulgarian EFL learners, and will retrieve and analyse verb forms produced by these two learner groups, before determining their function[18]. Therefore, the following chapter will only briefly outline the general methodology of the studies following the concept-oriented approach, together with their major results and pedagogical implications (see also Bardovi-Harlig 2007; 1999; Dietrich, Klein and Noyau 1995; Meisel 1987; von Stutterheim and Klein 1987 for a detailed outline of the concept-oriented research framework).

3.2. Concept-Oriented Approach to the Acquisition of Aspect

The concept-oriented approach to the second language acquisition of aspect is part of the functionalist approach to second language acquisition, which investigates the mapping of functions onto forms in interlanguage. It serves as a methodological framework of analysis of

---

[18] the omission of verb forms in contexts where they are grammatically "required" in the written language of Bulgarian and German learners will also be tackled, however in relation to other verb forms (see chapter 8)

interlanguage development, rather than as a linguistic theory or model itself (cf. Bardovi-Harlig 2007). The rationale behind the concept-oriented approach is the necessity to explain the psychological processes in second language acquisition apart from describing the linguistic products in L2 speech or writing. Meisel specifies that "we must define the concepts and functions which have to be encoded, and then analyze the devices used by different learners or types of learners to express these concepts and functions at different points on the developmental continuum" (Meisel 1987: 206).

Thus, the concept-oriented approach does not explore the development of aspect as a verbal category *per se*; it explores the development of temporal reference in interlanguage by analysing the range of linguistic devices used by second language learners to express the semantic concept of temporality, such as time adverbials, discourse-organisation patterns and verbal morphology. A basic argument for the concept-oriented approach in second language research is the assumption that "a second-language learner – in contrast to a child learning his first language – does not have to acquire the underlying concepts.[19] What he has to acquire is a specific way and a specific means of expressing them" (von Stutterheim and Klein 1987: 194). The main research objective of the concept-oriented studies is thus to identify the linguistic repertoire learners use to express temporality at a given stage in the acquisition process, as well as the development of this repertoire from one stage to another, including both targetlike and non-targetlike uses of temporal reference.

The majority of the concept-oriented studies are longitudinal case studies which rely on observations of untutored learners over a longer period of time and which involve either individual learners or small groups of learners (cf. Bardovi-Harlig 2007; Bardovi-Harlig 1999). In terms of data collection, the concept-oriented studies typically use elicitation measures such as personal narratives or interviews with the learners, as well as impersonal elicitation measures such as film or story-retelling tasks, in order to gain insights in the way learners encode temporality in narratives and other forms of dialogue (cf. Bardovi-Harlig 2007).

The findings of the concept-oriented studies reveal a consistent pattern in the development and use of temporal reference in interlanguage: namely, most studies agree on a specific order of emergence of the linguistic devices L2 learners employ in order to express temporality, which is in sharp contrast to the order of emergence of temporal markers in child

---

[19] unless these concepts are fundamentally different in L1 and L2

language acquisition (cf. Shirai 2009). Whereas children start using grammatical tense-aspect markers long before they start using lexical or other linguistic devices (cf. Shirai 2009: 169-170), beginning L2 learners lack verbal morphology altogether and start using it only after they have gone through all three stages of development in temporal reference: pragmatic → lexical → grammatical (cf. Shirai 2009; Bardovi-Harlig 2000).

In the initial stages of their language development, L2 learners resort to pragmatic means such as the so-called "scaffolded discourse" (cf. Meisel 1987: 212 – 213) or the chronological ordering of events in narrative discourse (cf. Schumann 1987). Meisel defines "scaffolded discourse" as a discourse in which beginners do not mark temporality explicitly, but receive clues from their native-speaker interlocutors, who help them by asking questions about events or happenings beyond the moment of speaking and often provide the answers themselves, waiting for the learner to point the relevant temporal reference (Meisel 1987: 212–213). Schumann (1987) confirms that in the very early stages of the acquisition process, learners do not mark verb forms for either tense or aspect, but relate temporal relationships by sequencing their utterances in the same order in which the reported events happen in real time (cf. Schumann 1987), i.e. by ordering them in such a way that "the order of mention follows the natural order" (Meisel 1987: 213).

In the later stages of the acquisition process, L2 learners begin to expand their linguistic repertoire by employing the first explicit linguistic devices such as deictic temporal adverbials (e.g. now, tomorrow, today), calendric expressions (dates, days of the week etc.), adverbs of time duration (e.g. always, three months) and coordinating conjunctions (cf. Schumann 1987; Bardovi-Harlig 1999; Noyau 2002). Notably, it is not until the very last stages of the acquisition process that learners start using morphological means such as verb inflections to encode temporality (Bardovi-Harlig 1999; Noyau 2002). Noyau (2002) attributes the order of emergence of formal temporal expressions to "the complexity of construction of [the] implied temporal notions" (Noyau 2002: 108), which determines the use of invariant forms such as single words expressing temporality in the early stages of acquisition and a gradual diversification of the linguistic devices in the later stages. Noyau's findings are in line with Andersen's (1984) "one-to-one principle" of early interlanguage construction – it specifies that "an IL system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form (or construction)" (Andersen 1984: 79). In contrast, in the later stages of the interlanguage construction, learners show a mixed output of verbal morphology and temporal adverbials,

relying on lexical means to a much greater extent than on verb inflections, which are still inconsistently supplied at that stage. As a result, learners show a completely different balance of morphological and lexical means expressing temporality than native speakers do, since verbal morphology in the target language "interacts with, supports, and often duplicates the work done by pragmatic devices in expressing temporality" (Schuman 1987: 38). Along similar lines, VanPatten (2002) explains second language learners' imbalanced output with their processing strategies which make them ignore the redundancy in temporal reference signalled by grammatical markers if the reference point has already been signalled by a lexical item such as a temporal adverb (cf. VanPatten 2002: 759-760).

The results of the concept-oriented studies suggest that the verb system of the target language is acquired comparatively late in the acquisition process, in contrast to other linguistic devices such as lexical means, which are processed and accessed by learners more easily – the result being a much earlier use. It has been argued that the reason for the late emergence of verbal morphology lies in the difficulty learners have processing grammatical forms carrying complex semantic information like temporal relations, as well as in the nature of grammaticalisation, which presupposes that learners can both identify the affixes or morphemes expressing temporality and map the correct functions onto them (cf. Gass and Selinker 1994; Bardovi-Harlig 1999; Noyau 2002). Grammaticalisation and the acquisition of complex syntactical structures do not emerge until the very advanced stages of the interlanguage development; in comparison, lexical acquisition happens all along the acquisition process. Moreover, even though temporal semantics is grammaticalised at that stage, the early second language use of verbal morphology still differs from the target-language use in terms of both frequency and accuracy of forms – a finding which has been addressed in greater detail by the form-oriented studies and which will be referred to in the next section. To illustrate, Dietrich, Klein and Noyau (1995) point out that the development of tense-aspect morphology in the morphological stage is slow and gradual and delineate a particular order in which tense-aspect markers occur in the acquisition process. At the beginning of the morphological stage, learners use a mixture of forms without necessarily differentiating between their functions (e.g. the base form V_0 alternates with V_{ing} in the same linguistic contexts), which leads to the conclusion that form precedes function in the early morphological stage. In the later substages of the morphological stage, V_{ing} and V_0 are slowly differentiated and their use is followed by the first irregular and regular past forms, the *be-going-to* future, the present perfect and the past perfect (Dietrich, Klein and Noyau 1995: 47).

Consequently, the concept-oriented studies demonstrate that learners follow a predictable path of development in the realisation of the concept "temporality" – from pragmatic to lexical to morphological means, irrespective of both L1 and L2. Thus, by identifying universal stages in the development of temporal reference, the concept-oriented studies focus on interlanguage as a linguistic system in its own right which develops independently from both the native and the target languages (cf. Bardovi-Harlig 2007). Nevertheless, one major factor determining the sequence of emergence of temporal expressions is learners' proficiency level – more advanced learners employ a more diversified linguistic repertoire, consisting of both verbal morphology and lexical means for the expression of temporality, whereas beginners use invariant verb forms and resort to pragmatic devices to compensate for the lack of morphological markers.

In spite of the limited number of learners involved, the concept-oriented studies offer valuable insights in the general patterns of development of temporal markers in interlanguage, since they focus on learners from a wide variety of L1 backgrounds acquiring different target languages. Their drawbacks lie primarily in the fact that the majority of the studies focus on beginners with little or no formal instructions, thus being able to explore the consecutive stages in the emergence of temporal semantics up to the morphological stage, but not the reasons for non-targetlike uses of verbal morphology in this final stage, or the possible effects of instruction on the development of tense-aspect morphology. Bardovi-Harlig (1995; 1997; 1997a) is one of the few researchers who examine the influence of instruction – she proves that even though formal instruction does not always guarantee success, instructed learners tend to outperform uninstructed learners in the long run in terms of the rates of the acquisition of tense-aspect morphology, as well as in terms of the formal accuracy of verb inflections. Nonetheless, she concludes that instruction only "change[s] the rate but not the route [my emphasis] of acquisition" (Bardovi-Harlig 1995: 164).

Hence, the concept-oriented approach offers a comprehensive, but fairly wide analytical framework for the exploration of the expressions of "time" in interlanguage and identifies a number of universal features in the development of second language temporal reference, irrespective of both the native and the target language of the learners; it can thus be applied to a variety of research contexts and second language studies. This being said, it is inadequate for the investigation of L2 morphology in the morphological stage of the L2 development and in particular for the investigation of a specific verbal category such as grammatical aspect.

3.3. Form-Oriented Approach to the Acquisition of Aspect

In contrast to the concept-oriented approach, the form-oriented approach (also form-to-function approach) to the acquisition of aspect adopts an entirely different perspective: its goal is to identify the presence of morphologically-marked verb forms in interlanguage and analyse their distribution prior to determining their function (cf. Bardovi-Harlig 1999). Similar to the concept-oriented approach, it does not focus on the acquisition of grammatical aspect only, but rather on the formal development of the entire verb system. Studies following the form-oriented approach are naturally concerned with the final stage in the development of temporal reference in interlanguage – the morphological stage. Whereas the early form-oriented studies investigate the emergence of tense-aspect morphemes somewhat accidentally, as part of a larger project on the first and second language acquisition of a wide range of inflectional morphemes in English, the later form-oriented studies focus exclusively on the analysis of emergent temporal inflections and aspect morphemes in particular, examining their distribution with respect to "higher-level constraints" (Robison 1995: 344) such as the lexical properties of verbs or the discourse structure of narratives. The next sections will review the early form-oriented studies – the morpheme order studies, as well as subsequent theory-driven studies testing the two major formal hypotheses concerning the second language acquisition of aspect: the Aspect Hypothesis and the Discourse Hypothesis.

3.3.1. The Morpheme Order Studies

Early examples of form-oriented studies exploring the emergence of tense-aspect morphology in interlanguage were the morpheme order studies from the 1970s and early 1980s, which aimed at predicting common patterns in the development of inflectional morphology in both first and second language acquisition. The morpheme order studies were mostly cross-sectional studies investigating the order of acquisition of up to 14 grammatical morphemes in English both as L1 and L2, ranging from articles and the third person singular –s to the –ing morpheme, the regular and irregular past morphemes (cf. Dulay 1974; Dulay and Burt 1974; Bailey 1974; Ellis 1994). Both first and second language acquisition studies examined grammatical morphemes used by children or learners with different L1 backgrounds and equated the accuracy with which children and second language learners used these morphemes with the order in which they acquired them, trying to establish a predictable order

of development, irrespective of factors such as the acquisitional type (L1 or L2 acquisition), age or the influence of the L1 of the learners in L2. Two types of methods were used to measure the accuracy and distribution of grammatical morphemes in child and learner language:

(1) an **obligatory occasion analysis**, which counted the obligatory contexts for a particular morpheme in the sample and measured the morphemes supplied correctly or incorrectly in these contexts

(2) a **targetlike-use analysis** (**TLU**), involving a count of the morphemes supplied correctly or incorrectly in obligatory contexts, as well as all morphemes overgeneralised to non-obligatory contexts (cf. Pica 1983).

The data elicited for the early morpheme order studies in first and second language acquisition research was mostly based on the Bilingual Syntax Measure (BSM) test, which tested children's knowledge of grammatical morphemes based on a picture description (cf. Dulay and Burt 1974; Gass 1994). The results showed a remarkable agreement in terms of a common hierarchical order in the acquisition of grammatical morphemes in English, which was assumed to reflect uniform developmental patterns in the acquisition of syntactic forms that both second language learners and children[20] went through.
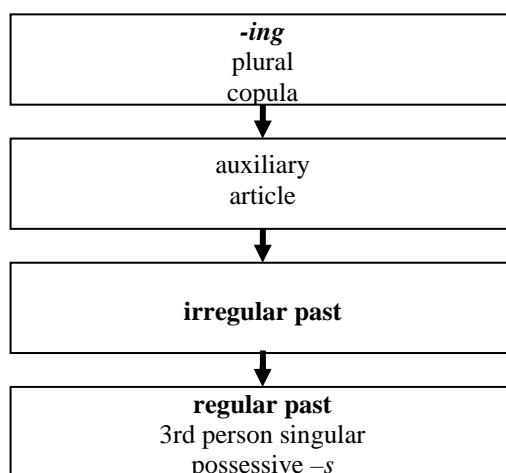
Figure 3.1. The order of acquisition of grammatical morphemes in English (adapted from Krashen 1977b in Krashen 1978: 190)

---

[20] Some of the morpheme-based studies focussing on adult L2 acquisition have found that formal L2 instruction clearly results in higher accuracy levels in comparison with the accuracy levels achieved by learners acquiring L2 morphemes in naturalistic environments (cf. Long 1983 and Norris and Ortega 2001 in Pica 2008: 3); however, they have rejected the idea that the order of introduction of morphemes in the classroom syllabus necessarily matches the order of L2 acquisition of morphemes – a theory postulated by Pienemann as the "teachability theory", according to which developmental stages cannot be influenced by formal teaching when L2 learners lack the "developmental readiness" to enter the next stage (cf. Pienemann 1987; 1999; Bardovi-Harlig 1995)

Although there was some variation between the results of the individual morpheme order studies, the majority agreed that there was a "more or less invariant order" (Gass and Selinker 2001: 131) or a so-called "natural order" in which grammatical morphemes were acquired (cf. Krashen 1978: 90) in both L1 and L2 acquisition. Nevertheless, since the accuracy levels of the individual morphemes were varied – some morphemes had very similar accuracy levels, whereas others remarkably different – the order of acquisition of individual morphemes had to be modified in subsequent studies. To this end, Krashen revised the "natural order", grouping morphemes in ranks (cf. Ellis 1994:78).

Krashen's "natural order" features several morphemes of the verb phrase encoding tense-aspect relationships: the progressive *–ing* morpheme, the copula, the regular and the irregular-past morphemes. To illustrate, the progressive *–ing* morpheme was defined as the first morpheme to be acquired by both children and L2 learners of English, typically followed by the irregular-past morpheme and lastly by the regular-past morpheme (cf. Bailey 1974; Krashen 1978). However, not all first and second acquisition studies obtained the same results in terms of the order of acquisition of the regular and irregular past: for instance, Dulay and Burt (1974) found out that the regular past preceded the irregular past in the speech of young second language learners. Likewise, contrary to Krashen's "natural order", the past progressive was found to succeed the simple past in interlanguage rather than preceding it, especially when both frequency and accuracy were taken into account (cf. Bardovi-Harlig 1999)

Possible explanations for the proposed "natural order" were offered by very few of the early morpheme order studies, mostly referring to potential reasons for early acquisition such as regularity, perceptual salience or complexity of the different morphemes. Thus, the progressive was defined as a perceptually salient or easily noticeable form, as well as formally stable and thus easily acquired by both children and L2 learners (cf. Bailey 1989; Noyau 2002; Bardovi-Harlig 1999).

Still, the majority of the morpheme order studies (and especially the early studies) offered insufficient explanations for the "natural order" of morpheme acquisition. To illustrate, Goldschneider and DeKeyser criticise the "very little concern [of the morpheme order studies] about explaining the order findings" (Goldschneider and DeKeyser 2001: 11) and suggest a combination of five factors which influence the results obtained for both first and second language acquisition of grammatical morphemes: perceptual salience, semantic

complexity, morphophonological regularity, syntactic category and frequency in the input (Goldschneider and DeKeyser 2001: 11). They regard most of these factors as variants of salience, arguing that a morphological one-to-one relationship between form and meaning is more salient and easier to acquire than varying forms and allomorphy. Furthermore, they claim that free morphemes are more noticeable than bound morphemes, just as high-frequency morphemes and lexical morphemes are more salient than lower-frequency morphemes and grammatical morphemes (Goldschneider and DeKeyser 2001: 35-36). Along similar lines, Bardovi-Harlig (2000) specifies that a variety of factors such as "semantic complexity, syntactic complexity, frequency of input and functional load are all likely to contribute to determining the acquisition order and should be further investigated in second language acquisition [research] (Bardovi-Harlig 2000: 181). Likewise, several recent studies like Ellis and Larsen-Freeman (2006), Ellis (2008) and Ellis and Collins (2009) conclude that only a combination of multiple variables in the input like construction frequency, type and token frequency, salience and perception, as well as form can influence and explain "a substantial 71 per cent of the variance in the acquisition order" (Ellis and Larsen-Freeman 2006: 559). In particular, Ellis and Collins (2009) emphasize the frequency of verb types and their distribution in the input as influencing productivity in the second language, together with the salience of temporal verb inflections, which they identify as low in comparison to the salience of time adverbials (cf. Ellis and Collins 2009: 330-331).

To resume, apart from their inability to explain the order of emergence of grammatical morphemes, the early morpheme order studies have become an object of severe criticism over the past twenty years due to several other drawbacks. A major drawback of the early morpheme order studies is their preoccupation with the accuracy of surface forms and their total neglect for the appropriate use of these forms. Most of the early morpheme order studies dealing with the L2 acquisition of morphemes have ignored the fact that acquisition of form does not necessarily involve acquisition of meaning, or "grammatical well-formedness and appropriate use of forms do not develop simultaneously" (Bardovi-Harlig 1992: 253). Bardovi-Harlig has shown that the systematic use of tense-aspect markers in interlanguage precedes targetlike use, pointing out that even intermediate learners demonstrate a high rate of non-targetlike associations between form and meaning, making much more errors in the use than in the form of the tense-aspect morphemes (cf. Bardovi-Harlig 1992; Bardovi-Harlig 1999). Another point of criticism concerns the methodological problem of linking the accuracy levels of individual morphemes to the exact order of their acquisition, a point which has been revised to a certain extent by the introduction of group morpheme ranks instead of

individual morpheme ranks, but which is still heavily critised as misrepresentation of an actual order of difficulty (rather than order of acquisition). A third methodological weakness concerns the use of an obligatory occasion analysis as an accuracy measurement, which considers only obligatory contexts and fails to account for oversuppliance of grammatical morphemes in non-obligatory ("wrong") contexts. This method of calculation has been successfully replaced by a targetlike-use (TLU) analysis – an accuracy measurement which considers morpheme use in both obligatory and in non-obligatory contexts (cf. Pica 1983; Ellis 1994). A fourth problem relates to the results obtained by the early morpheme order studies – more recent corpus-based studies like Tono (2000) and McEnery et al. (2006) reveal a different order of acquisition of morphemes that only partially overlaps with Krashen's "natural order", thus arguing that the proposed universal morpheme order does not hold for all types of data (cf. Tono 2000: 131-132; McEnery et al. 2006: 262). Most recently, the "natural order" in L2 morpheme acquisition has been severly challenged by two learner corpus studies which have demonstrated that the lack of a morpheme in a particular L1 clearly results in a lower accuracy order in L2 in contrast to L2 learners whose L1s mark this particular morpheme (e.g. articles), thus proving a much bigger importance of L1 influence in L2 morpheme acquisition than previously found (cf. Murakami 2013: 325).

Despite their limitations, the morpheme order studies are the first studies to reveal general patterns in the formal development of syntactic structures, which are common to learners of English from various L1 backgrounds, acquisitional environments and age groups. Although these patterns may not be rigid, they illustrate universal trends in the development of inflectional morphology in English as a first and second language irrespective of the L1 influence or even the effects of instruction. Therefore, the early morpheme order studies serve as evidence to support a weak version of the developmental pattern hypothesis proposed by Rod Ellis (1994: 111), which claims that "[w]eaker evidence [for universal developmental patterns] is found if it is shown that an order or a sequence [of acquisition] applies only to specific L2s and/or to specific groups of learners" (Ellis 1994: 111).

3.3.2. Recent Perspectives to the Form-Oriented Approach

A number of more recent form-oriented studies focus on the general patterns of second language development of the English verb system and the order of acquisition of tense-aspect morphemes in particular (Bardovi-Harlig 2000; Dietrich 1995; Housen 2000; Housen 2002a; Housen 2002b). The majority of these studies delineate three major stages in the formal

acquisition of tense-aspect morphemes: an initial stage that lacks verb forms altogether or features them as unanalysed "formulaic expressions (e.g. *I don't know*)" (Housen 2002b: 157); a second stage which involves the use of unmarked verb forms and the occasional inflected verb form such as the *–ing* participle or the irregular past tense of frequent lexical verbs, functioning as default forms; and a third stage which involves a gradual differentiation between the individual verb forms (cf. Dietrich 1995; Housen 2002a; 2002b). In keeping with Bardovi-Harlig's findings (cf. Bardovi-Harlig 1992; 1999; 2000 etc.), Dietrich, Klein and Noyau (1995) in their large-scale study on the acquisition of temporality by 21 learners with different L1 backgrounds acquiring a number of target languages (English and several other languages – mostly of Germanic origin) observe that the emergence of verb forms in interlanguage precedes a proper differentiation between their functions. The results of their study show that base forms continue to coexist with inflected forms in the same linguistic contexts even in the later stages of the acquisition process, confirming thus Bardovi-Harlig's claim that "form-meaning associations are not complete until the entire system is complete" (Bardovi-Harlig 2000: 175). The three general stages in the formal development of second language temporal reference (irrespective of both the target and source languages) have been summarised by Dietrich, Klein and Noyau (1995: 264-271) and are presented in an adapted version in table 3.1:

Stage	Properties
1. Pre-basic varieties	• lexical stage – bare verbs • no functional inflections • "principle of natural order"
2. Basic varieties	• learner repertoire of adverbials increases • lexical verbs in the base form
3. Post-basic varieties	• co-existence of different morphological verb forms • no proper functional differentiation • no distinct developmental stages • tense marking precedes aspect marking • irregular morphology precedes regular morphology

Table 3.1. Stages in the L2 development of temporal reference (adapted from Dietrich et al. 1995: 64-271)

A more detailed outline of the order of emergence of formal verb categories in L2 English is given by Housen (2002b) on the basis of data gained from a longitudinal study of Dutch and French EFL learners:

Stage	Category	Comment	Example
0	Invariant verb form	V_0	*see, play*
1	Present participle *Ving* Irregular past of be	Initially without aux. *be*	*seeing, playing* *was*
2	Irregular past of other verbs		*had, got*
3	Regular past *Ved* Future be-going-to + *Vinf*	Allomorphs: without aux. *be, to, -ing, gonna*	*played, worked* *is going married* *are go dancing* *am going to take* *is gonna happen*
4	Perfect aux + *V* Present *Vs* Future will + *V*	Allomorphs: aux. *be* and *have*; initially $V = V_0$	*have see* *is fall* *is fallen* *has fall* *have fallen* *goes, comes,* *will make, will see*

Table 3.2. Order of emergence of tense-aspect forms (adapted from Housen 2002b: 158)

Both Dietrich et al. (1995) and Housen (2002b) agree on the fact that the first aspect form to appear in L2 English is the progressive *–ing* form, which is also the first compound form used by EFL learners. Initially, it is realised by the bare progressive (V_0), before grammatical marking expands to include the auxiliary *be* too: first in its present form (*is, are*) and then in its past form (*was, were*). In contrast, the second aspect construction in English – the perfect – has been found to emerge significantly late in interlanguage, since both the present and the past perfect depend on "the stability of the simple past" (Bardovi-Harlig 2000: 149), which is also acquired relatively late.

In terms of the functional/conceptual differentiation between tense-aspect morphemes, Housen (2002b) states that the first concept which is systematically marked in interlanguage is the concept "anteriority", variably expressed by the simple past or by the present perfect, followed by the grammatical encoding of imperfectivity/progressivity, marked with the auxiliary *be* + *Ving*. Further temporal concepts like futurity and habituality (auxiliary *be* + *going* + *Ving*, simple present V_0/Vs) are marked later on in the acquisition process; similarly, the differentiation between the simple past as a deictic past and the present and past perfects as anaphoric pasts appears in the last stages of the acquisition process (Housen 2002b: 162-163). The order of functional development is summarised in table 3.3.

Stage	Meaning	Form
0	pre-functional stage	Random and complementary distribution
1	anteriority	1. Past of *Be* (*was*) 2. Irreg. Past (other verbs) 3. Perfect *Have/Be + V* 4. Reg. Past *Ved*
	imperfectivity/progressivity	1. Aux. *Be + Ving*
2	futurity	1. Aux. *Be + going + Vinf* 2. Aux. *Be + Ving* 3. Aux. *Will + V* 4. Present V_0/Vs
	habituality	1. Aux. *Be + Ving*
	present	1. Present V_0/Vs
3	simple past	1. Past of *Be* (*was, were*) 2. Irreg. past 3. Reg. past *Ved*
	present perfect	1. *Have/Has + Ved/Virreg*
	past perfect	1. *Had + Ved/Virreg*

Table 3.3. Order of functional marking of tense-aspect meanings and their respective markers (Housen 2002a: 162)

The formal and functional development of tense-aspect morphemes in English as a second language proposed by Housen (2002b) seems to confirm the majority of the findings concerning developmental stages in L1 and L2 morphology which were put forward by the early morpheme order studies. However, Housen (2002b: 164-165) emphasises the fact that individual variation may also be at play, influencing the acquisition order and the preference of one form over another. Nevertheless, there is an agreement between the results of the early and the late form-oriented studies with respect to the formal and functional acquisition of the progressive and the perfect as the two aspect constructions in English: studies from both research strands specify that the progressive aspect is acquired first – both formally and functionally; whereas the perfect aspect is acquired much later and in interdependence with the acquisition and accuracy rates of the simple past (Housen 2002b; Bardovi-Harlig 1997; 2000). Housen (2002b: 164-165) tentatively suggests that the reason for the late acquisition of the perfect aspect lies in its non-deictic nature and its relative markedness in comparison to deictic past tenses, which are more common in the world's languages and presumably easier to acquire (cf. Bybee 1985: 160-161); (Housen 2002b: 164-165).

To sum up, the formal and functional development of the second language verb system is a long and gradual process which "resembles the slow mastering of a skill, such as piano playing, much more than an increase of knowledge, such as the learning of a mathematical

formula" (Dietrich, Klein and Noyau 1995: 270). Notably, the emergence of one verb form affects all other verb forms in the second language verb system in terms of the forms' overall appropriate use, since an overgeneralization of one verb form (e.g. the present perfect) in the contexts of another verb form (e.g. the simple past) leads to lower rates of accuracy for both forms (cf. Bardovi-Harlig 1997). Inappropriate uses of tense-aspect morphemes in the morphological stage of the temporal development in interlanguage have been further investigated with respect to the influence of other factors such as the properties of lexical verbs or the patterns of discourse organisation in narratives. These factors have resulted in the postulation of two major hypotheses about the acquisition and use of tense-aspect morphology – the Aspect Hypothesis and the Discourse Hypothesis – which will be dealt with in more detail in the next two sections.

3.3.3. The Aspect Hypothesis

The first influential hypothesis concerned with the emergence of verbal morphology is the Aspect Hypothesis, which is based on the assumption that the acquisition and development of verb inflections is influenced by the semantic properties of verbs these inflections are attached to – i.e. learners tend to select verb inflections according to the verbs' inherent lexical aspect. This hypothesis has been postulated in two different versions: an early or strong version which has been variably called the "Defective Tense Hypothesis" (DTH) (cf. Weist 1984) or the "Aspect-Before-Tense Hypothesis" (Bloom et al. 1980 in Shirai 2009: 173) and which focuses exclusively on verbal morphology in L1 acquisition, and a more recent version focussing on L2 acquisition which has been called the "Primacy of Aspect Hypothesis" (POA) (cf. Robison 1990) or simply the "Aspect Hypothesis" (cf. Shirai 1995). Both the early and the late versions have their origins in studies on child language acquisition and rely on the universalist idea that in the early stages of first language acquisition, children do not code either tense or grammatical aspect, but use verbal morphology to mark redundantly the inherent lexical aspect of verbs, irrespective of the temporal context these verbs occur in. Thus, the Defective Tense Hypothesis replicates previous studies which claim that children have a cognitive deficit expressing deictic tense linguistically and that they can only refer to immediate past events with a clear end or result, i.e. children are only able to recognise the inherent aspectual qualities of verbs in the early stages of first language acquisition (cf. Weist 1984; Andersen 1996).

Derived from the cognitive-deficit claim, the DTH in its absolute form argues that "emerging tense morphology is defective in its function since it does not code deictic relationships" (Weist 1984: 348); i.e. children mark only telic verbs with past-time inflections, refer only to immediate past situations and make tense distinctions only if they are accompanied by inherent aspectual distinctions, marking them incorrectly and redundantly (Weist 1984: 348). In a similar vein, the Primacy of Aspect Hypothesis (cf. Robison 1990) postulates for SLA that "aspect is primary in the sense *not* that morphemes that denote aspect in the target language are acquired first, but that target language verbal morphemes, independent of their function in the target language are first used by the learner to mark [lexical] aspect" (Robison 1990: 316, my emphasis). Both versions of the Aspect Hypothesis define the inherent lexical aspect of verb phrases in terms of Vendler's fourfold classification into states, activities, accomplishments and achievements (see chapter 2.1.1.); both hypotheses are concerned with the relationship between the lexical verb types and their grammatical markings. However, they differ in the extent to which they characterize the interdependence between lexical and grammatical aspect in child language and in interlanguage.

Thus, the Defective Tense Hypothesis has been criticised for its rigid correlation between verb inflections and inherent lexical aspect, as well as on account of the cognitive-deficiency claim concerning children's underdeveloped concept of temporality (cf. Andersen 1996; Weist 1984). A further point of criticism concerns the related idea that *both* lexical and grammatical aspect are acquired before tense, which has been suggested by several studies in child language acquisition, disregarding the lexical-grammatical aspect distinction (for a detailed review, see Andersen and Shirai 1996). This idea rests on a typological observation which specifies that more of the world's languages have aspect markings than tense markings (both inflectional and derivational), and that aspect markings in the world's languages are closer to the verb stem than tense markings are, thus contributing to a greater extent to the intrinsic meaning of verbs (cf. Bybee 1985: 29-31). However, evidence from languages grammaticalising the perfective-imperfective distinction contradicts this view by proving that children acquire the imperfective inflections fairly late (cf. Weist 1984, Andersen and Shirai 1996). Likewise, both longitudinal (e.g. Dietrich, Klein and Noyau 1995) and cross-sectional (e.g. Housen 2002a; 2002b) studies investigating tense-aspect morphology in English and other target languages have confirmed that grammatical aspect markings are acquired *after* tense markings, or "tense marking precedes aspect marking" (Dietrich, Klein and Noyau 1995: 47, see also table 3.3). Consequently, a Relative Defective Tense Hypothesis has been

adopted instead of the Absolute Defective Tense Hypothesis in first language acquisition, specifying again interdependence between inherent lexical aspect and grammatical tense-aspect markings in child language, but excluding the cognitive-deficiency claim (cf. Andersen 1996). In SLA research, the most widely-used hypothesis is the Primacy of Aspect Hypothesis (POA) or simply the Aspect Hypothesis, which has been reformulated to include four separate claims concerning the relationship between inherent lexical aspect and grammatical markings:

(1) Children first use past marking (e.g. English) or perfective marking on achievements and accomplishment verbs, eventually extending its use to activity and stative verbs.
(2) In languages that encode the perfective – imperfective distinction, imperfective past appears later than the perfective past, and imperfective past marking begins with stative verbs and activity verbs, then extending to accomplishment and achievement verbs.
(3) In languages with progressive aspect, progressive marking begins with activity verbs, then extends to accomplishments and achievement verbs.
(4) Progressive markings are not incorrectly overextended to stative verbs. (Andersen and Shirai 1996: 533)

The underlying assumption of the Aspect Hypothesis is that verb inflections in interlanguage are strongly influenced by the semantic properties of the verbs they are attached to; consequently, verb inflections occur in a non-targetlike complementary distribution, so that accomplishment and achievement verbs are predominantly marked for the perfective past, whereas state verbs are marked for the imperfective past and activities for the progressive (cf. Bardovi-Harlig 1999; 2000). In contrast, native-speaker distribution of verb inflections has been claimed to be less biased, allowing for contrasts between verb inflections attached to the same type of lexical verb, i.e. stative verbs in native-speaker production can receive all three of the above-mentioned tense-aspect inflections, depending on the immediate linguistic and discourse context. The Aspect Hypothesis has been formally summarised by Housen (2000), who matched the distribution of the progressive, perfective and perfect formal markings across Vendler's four lexical verb types with the developmental stages in the acquisition of English tense-aspect morphology (cf. Housen 2000: 250).

Progressive aspect morphology (*-ing*)				Past-Perfect tense morphology (*-ed/en*)			
STA	ACT	ACC	ACH	STA	ACT	ACC	ACH
stage							
0	V	V	V	V	V	V	V	
1	V	V-ing	V	V	V	V	V	V-P
2	V	V-ing	V-ing	V	V	V	V-P	V-P
3	V	V-ing	V-ing	V-ing	V	V-P	V-P	V-P
4	(V-ing)	V-ing	V-ing	V-ing	V-P	V-P	V-P	V-P

Table 3.4. Development of English Tense-Aspect Morphology (Housen 2000: 250)

The four claims of the Aspect Hypothesis have been tested in a variety of L2 acquisition studies exploring the acquisition of several major target languages (English, Dutch, French, Japanese, Italian and Spanish) from the perspective of a wide range of L1 backgrounds (cf. Bardovi-Harlig 1995; Collins 2002; Collins 2004; Robison 1995; Rohde 2002; Shirai 2007; Sugaya 2007; Robison 1990). The majority of these studies focus on English as the target language of adult second language learners (with the exception of Rohde 2002 who focussed on children) with various L1 backgrounds, some of whom acquiring English in a naturalistic environment, others in instructed environments such as ESL intensive courses (for a comprehensive review of the Aspect Hypothesis studies, see Andersen and Shirai 1996; Bardovi-Harlig 1999; 2000). Various tasks have been employed to elicit L2 data – mostly in experimental settings, involving conversational interviews and film-retelling tasks or written exercises such as cloze passages eliciting the correct verb inflection. Coding has mostly followed Vendler's four-way distinction for inherent lexical aspect, with the exception of Robison (1995) who employed the six-way distinction illustrated in section 2.1.1. Verbs have been characterised into Vendler's lexical aspect categories with the help of various diagnostic tests, mostly following Andersen and Shirai's (1995) test model which helps to differentiate between states and non-states, activities and non-activities, and accomplishments and achievements (cf. Andersen and Shirai 1995). Most of the studies have had a cross-sectional design, involving L2 learners from various proficiency levels up to the upper-intermediate level, whereby proficiency levels have been variably measured with the help of placement tests or calculation of the years of exposure to English in the host country or the years of formal English instruction (cf. Bardovi-Harlig 1999; 2000).

The methods of quantification used in the studies testing the Aspect Hypothesis answer two different questions:

(1) "Where do verb inflections occur?"
(2) "How are aspectual categories marked?" (Bardovi-Harlig 2002: 133-135)

These two questions give way to two different types of analysis: across-category analysis and within-category analysis (cf. Bardovi-Harlig 2000; 2002). The across-category analysis counts the distribution of verb inflections across Vendler's four aspectual classes, e.g. calculating the percentage of all progressives that are activities. The within-category analysis calculates the percentage of all activities that are marked for the progressive aspect. These two different methods of quantification yield different results, e.g. 80% of the progressives in a sample of data may be activities, whereas 25% of the activity verbs in the same sample of data may be marked for the progressive aspect. The within-category analysis has been more widely used, since it is not as sensitive to an unequal distribution of lexical verb tokens across Vendler's four aspectual categories as the across-category analysis is, i.e. there have usually been more telic than atelic verbs in the samples used by the studies testing the Aspect Hypothesis (cf. Bardovi-Harlig 2000; 2002). However, since the focus of the present study lies on the realisation of two particular aspect forms in advanced L2 English – the progressive and the perfect – an across-category analysis lends itself better for a quantification of e.g. the frequencies of the progressives which are states, activities, accomplishments or achievements.

The findings of the studies testing the Aspect Hypothesis in English as a target language are unanimous with regard to the first claim of the Aspect Hypothesis, i.e. the fact that perfective past markings are readily and consistently used by learners with telic verbs such as achievements and accomplishments and slowly expand to state and activity verbs in the interlanguage development (cf. Bardovi-Harlig 1995; Collins 2002; 2004). Bardovi-Harlig (2000) explains this particular finding by referring to the nature of elicitation tasks such as narratives which presuppose a high number of accomplishment and achievement verbs, pointing out that "the perfective past is the first past morpheme acquired and thus easily observed in the interlanguage of learners who have reached the morphological stage of temporal expression" (Bardovi-Harlig 2000: 228). A related hypothesis termed the "Default Past Tense Hypothesis" was put forward by Salaberry (1999), who claimed for the second language acquisition of Spanish that learners use the "Spanish preterite […] as a default

marker of past tense during the beginning stages of L2 acquisition" (Salaberry 1999: 171), irrespective of the inherent lexical aspect of verbs. Other elicitation methods such as cloze tests in English (Bardovi-Harlig and Reynolds 1995; Collins 2002; 2004) confirm the distribution of perfective past inflections with telic verbs. Very few studies contradict the first claim of the Aspect Hypothesis, one of them being Rohde's (1996; 2002) longitudinal study on German children acquiring English as a second language in a naturalistic environment. On the basis of the children's oral production, Rohde observes that a large number of achievement verbs in their speech remain uninflected, whereas almost all state and activity verbs are inflected. Rohde offers possible explanations of this unusual distribution referring to specific learner- and task-related variables such as the L1-L2 combination and the conditions of naturalistic second language acquisition (cf. Rohde 2002: 216-217).

The second claim of the Aspect Hypothesis – the spread of the imperfective past after the perfective past in languages encoding the perfective-imperfective distinction – has also been confirmed by the majority of the studies investigating second languages such as French or Spanish, thus corroborating the results obtained for first language acquisition (cf. Weist 1984; Bardovi-Harlig 1999; 2000). Likewise, the third claim of the Aspect Hypothesis relating to the gradual spread of the progressive from activities to accomplishments, achievements and states has been proven by a number of studies investigating learner use of the progressive in English as a second language (e.g. Collins 2002; 2004; Robison 1995; Bardovi-Harlig and Reynolds 1995; Bardovi-Harlig 2000). To illustrate, Collins (2002) and Bardovi-Harlig (2000) observe that most of the learners in their sample use the progressive with activity verbs, subconsciously responding to "the durativity of activities in their use of progressive forms by marking lexical aspect redundantly with morphological aspect" (Bardovi-Harlig 2000: 238). The percentage of activity verbs marked for the progressive gradually drops with increasing proficiency; moreover, the progressive develops formally from a bare progressive form to a tensed progressive featuring the inflected auxiliary (cf. Bardovi-Harlig and Reynolds 1995; Bardovi-Harlig 2000, Housen 2000; 2002 etc.). Still, the third claim of the Aspect Hypothesis has only been partially confirmed by studies focussing on other target languages – Sugaya and Shirai (2007) argue that in the case of the progressive in L2 Japanese, lower-level learners show little preference for activity verbs when using the Japanese progressive marker.

The fourth and last claim of the Aspect Hypothesis – the lack of overgeneralization of the progressive to stative verbs – has stirred severe disagreements among researchers, since some

of the studies have found overextended use of the progressive with stative verbs (e.g. Robison 1990), while others have found none (e.g. Bardovi-Harlig and Reynolds 1995; Rohde 2002; Collins 2002). Typological studies comparing the progressive across the world's languages support the fourth claim by stating that "progressives are preferred with dynamic verbs – either activity or process verbs" (Bybee 1994: 139). Nevertheless, L2 studies proving overgeneralisation of the progressive to stative contexts have emphasised the fact that L2 learners with L1s encoding the perfective-imperfective distinction may try to find parallels between the imperfective aspect in their native language and the progressive in English, since progressivity is part of imperfectivity (cf. Comrie 1976; Andersen and Shirai 1996, see also section 2.1.). Therefore, the L1 influence is regarded as one of the many factors which influence the overextension of the progressive to stative verbs, the others being universal factors which interact with the individual, learner- and learning-related factors in the L2 acquisition of tense-aspect morphology. Hence, Shirai (2009) advocates that "[f]uture research should systematically investigate the effect of the L1 by comparing different L1 groups acquiring the same language to tease out the effect of natural acquisitional processes from the effect of L1 […] (Shirai 2009: 184).

Even though the individual claims of the Aspect Hypothesis and in particular the fourth claim have not been confirmed by all studies, most studies agree on the fact that whereas "the exact pattern [of distribution of verbal morphology] will vary depending on L1, L2, and individual differences between learners, verbal morphology correlates with lexical aspect at least during some stage in the development of an interlanguage" (Robison 1990: 330). Therefore, the general claim that tense-aspect morphology in the morphological stage of interlanguage development depends on the inherent lexical aspect of verbs seems to be validated by the vast majority of the studies testing the distribution of tense-aspect morphemes across different target and native languages, acquisitional environments and elicitation tasks.

Several explanations for the interdependence between lexical and grammatical aspect have been offered by researchers, ranging from a distributional bias in the input learners receive to a prototype theory of the most prototypical carriers of a particular tense-aspect morpheme. Both explanation strands have been summarised by Andersen and Shirai (1995; 1996). The distributional bias hypothesis specifies that "N[ative]S[peaker]s' speech to nonnative speakers or to young L1 learners exhibits a distributional bias [of morphological forms] that matches the POA predictions much more closely than in the speech to other NSs"

(Andersen and Shirai 1996: 560). In other words, native speakers of English show a distributional bias of verbal morphology in their speech, especially when it is directed to children and non-native speakers, so that activity verbs are often marked for the progressive and accomplishments and achievements for the perfective past. As a result, learner language simply mirrors and enhances a similar, though weaker distributional bias in native speakers' speech. The second explanation – the prototype account – has its origin in studies on prototype semantics (cf. Shirai and Andersen 1995: 757-758), which claims that each category has its most prototypical member serving as its best example. In terms of tense-aspect morphology, both children and learners select the most prototypical meaning of tense-aspect morphemes and attach them initially to the most suitable members of each verb category, gradually expanding the use of tense-aspect markings to other, less prototypical verbs. As a result, learners select the most prototypical meaning of the perfective past – a single punctual event with an end or a result – and attach the perfective past marker to the best-suited achievement or accomplishment verbs (e.g. *fall, drop*), before they start using the perfective past inflection with other verbs, such as activities or states (cf. Andersen and Shirai 1995; 1996).

Undoubtedly, the Aspect Hypothesis has established itself as a major theoretical framework for the analysis of the interdependence between inherent lexical aspect and the distribution of grammatical tense-aspect markings in the early and intermediate stages of the second language acquisition process. Despite the controversies between the individual studies in terms of methodology, research design, quantification and learner-related variables, the majority of the studies have confirmed the first three claims of the Aspect Hypothesis, documenting a strong interdependence between Vendler's verb classes and learners' selection of tense-aspect inflections, especially in the initial stages of the language acquisition process. Another widely-used theory-driven framework for the analysis of the distribution of tense-aspect morphemes in interlanguage is the Discourse Hypothesis, which assumes a correlation between tense-aspect morphemes and the discourse organisation of narratives. The Discourse Hypothesis will be dealt with in the next section.

3.3.4. The Discourse Hypothesis

The Discourse Hypothesis is the second major form-oriented theoretical framework for the analysis of tense-aspect morphemes in interlanguage – it investigates their distribution with respect to the discourse structure of narratives. Its origins lie in Hopper's (1989) cross-

linguistic study on the relationship between aspect and discourse, which specifies that aspectual distinctions depend on the structure of narrative discourse and more specifically on the foreground – background distinction in narratives. Hopper uses Dry's (1983) characterisation of narratives as structures of consecutive events which "create for the reader or hearer an imagined timestream as a dimension of the narrative world in which the events occur" (Dry 1983: 19). Moreover, all narratives as defined by Hopper (1989) consist of an "actual story line and the language of supportive material which does not itself narrate the main events" (Hopper 1989: 213). The actual story line is referred to as "the foreground" or the backbone of the narrative, whereas the supportive material is referred to as "the background" of the narrative (cf. Hopper 1989; Boogaart 2007; Dry 1983; Bardovi-Harlig 1994; Bardovi-Harlig 1999). The foreground and the background of a narrative differ from each other with respect to the events they relate and to the sequencing of these events: thus events from the main storyline are in the foreground of a narrative and are set in a sequential order mirroring their actual order of happening in real time; in contrast, "shunted" events (Hopper 1989: 214) are in the background of a narrative, not sequenced with the foregrounded events and often simultaneous with them (cf. Hopper 1989: 213-214). Thus, foregrounded clauses move forward the narration, whereas backgrounded clauses only comment, explain or enhance the main narrative line. Dry offers a more precise definition of foregrounded clauses, arguing that foregrounded clauses "refer to a point on the timeline and trigger a perception of time movement" (Dry 1983: 48), whereas backgrounded clauses do not refer to a single point in time or refer to a point that has already been introduced as part of the given information (Dry 1983: 32-33).

The foreground-background distinction is particularly useful for understanding the distribution of grammatical tense-aspect markings: thus in languages coding the perfective – imperfective distinction, perfective aspect is found primarily in foregrounded clauses, rendering "dynamic, kinetic events" (Hopper 1989: 216), whereas imperfective aspect is found in backgrounded clauses and descriptive situations (the main properties of the perfective-imperfective distinction with respect to narrative discourse types are summarised in table 3.5). In addition, in terms of the relationship between discourse structure and the inherent lexical aspect of verbs, Dry argues that "accomplishment and achievement sentences always move time [forward], and activity and stative sentences may move time if it is clear from the context that the situation represented in the sentence is the outcome of a change of state" (Dry 1983: 23). Similarly, Hopper (1989) characterises the correlation between grammatical aspect, inherent lexical aspect and discourse structure as "a tendency for

punctual verbs to have perfective aspect and to occur in foregrounded sentences, and conversely for verbs of the durative/stative/iterative verb types to occur in imperfective, i.e. backgrounded clauses" (Hopper 1989: 215).

PERFECTIVE	IMPERFECTIVE
• foregrounding • event indispensible to narrative	• backgrounding • state or situation necessary for understanding motives, attitudes etc.
• chronological sequences • view of event as a whole, whose completion is a necessary prerequisite to a subsequent event	• simultaneity or chronological overlapping of situation C with the event A and/or B • view of a situation or happening whose completion is not a necessary prerequisite to a subsequent happening
• human topics	• variety of topics
• dynamic, kinetic events	• statis, descriptive situations

Table 3.5. Properties of the perfective-imperfective aspect distinction (adapted from Hopper 1989: 216)

Like the Aspect Hypothesis, the Discourse Hypothesis has not remained unchallenged: Hopper's account of the distribution of inherent lexical verb types serving to distinguish between foreground and background in discourse has been criticised as oversimplified and ignoring the fact that "both durative *and* punctual verbs will as a rule be interpreted as backgrounding their events if they occur in *subordinate* temporal clauses" (Couper-Kuhlen 1994: 231, original emphasis). Furthermore, Couper-Kuhlen (1994) in her study on American conversational narratives shows that the distribution of grammatical aspect markers – and in particular the progressive aspect – does not follow the Discourse Hypothesis, but appears to be "in flagrant violation of standard rules for aspectual use in English narration" (Couper-Kuhlen 1994: 229), since progressive markers in her sample often occur in foregrounded clauses. Nevertheless, this last finding can also be explained with the nature of her sample (American conversational narratives) and the general preference of American English (and American English conversation in particular) for the progressive, as well as with the fact that progressives occur mostly in main clauses (see also section 2.2.1).

Still, studies testing the Discourse Hypothesis in a second language acquisition context have demonstrated a clear correlation between discourse structure and learners' choice of tense-aspect morphemes in written and oral narratives (cf. Bardovi-Harlig 1995; Bailey 1989; Jappy 1996). In a sample testing ESL learners of different proficiency levels, Bailey (1989) found out that learners related the past progressive with backgrounded clauses, which they preferred to use before the foregrounded clauses featuring the simple past, reflecting thus a

chronological order in which background or old information preceded foreground or new information. Along similar lines, Bardovi-Harlig (1995) studied a group of ESL learners of different proficiency levels and established a strong correlation between the use of the simple past of lexical verbs and the foreground of narratives; moreover, she found a greater morphological diversity in the background clauses, which she classified as "promising for the study of perfect and progressive" (Bardovi-Harlig 1995: 285). Both studies emphasized the role of the level of proficiency of learners, which determined the strength of the relationship between grounding and tense-aspect morphemes – the higher the proficiency of learners, the weaker the correlation between grounding and tense or aspect.

The Discourse Hypothesis explains the distribution of tense-aspect morphemes with respect to narrative discourse as the most suitable type of discourse which consists of a chronological ordering of events; however, research has shown that other types of discourse like description also influence second-language learners' selection of tense-aspect inflections. Notably, it has been shown that lower-level learners tend to rely on descriptions presupposing the use of the present tense, so that their avoidance of tense-aspect markings resembles a continuity of tenses which "give[s] the appearance of being deliberate present tense continuity" (Godfrey 1980: 108). Admittedly, different types of narrative elicit different proportions of foreground and background material – for instance, personal narratives have been reported to elicit the richest foreground and background information (cf. Bardovi-Harlig 1999). In addition, Slabakova (2002) suggests that the Discourse Hypothesis is more "useful in accounting for the production patterns of advanced learners" (Slabakova 2002: 180) in terms of tense-aspect morphology, rather than those of beginners or intermediate learners, who are the preferred respondents in studies testing the Aspect Hypothesis.

Although most studies exploring the second language acquisition of aspect have adopted either the Aspect Hypothesis or the Discourse Hypothesis as a form-oriented theoretical framework, researchers have observed a considerable overlap between the two hypotheses with respect to the interdependence between tense-aspect markings, inherent lexical aspect and discourse structure. For example, the Aspect Hypothesis claims that telic verbs such as accomplishments and achievements will carry past-tense or perfective morphology, and the Discourse Hypothesis predicts that telic verbs will occupy the foreground of a narrative; therefore, whenever telic verbs are marked with simple past morphology and appear in the foreground, the two hypotheses overlap. They overlap too whenever atelic verbs appear in the background and lack perfective markings – in these two

cases the Aspect and the Discourse hypotheses are difficult to keep apart (cf. Bardovi-Harlig 1994: 55). However, whenever telic verbs are in the background of a narrative and atelic verbs in the foreground (e.g. Dry 1983), the two hypotheses differ from each other. Bardovi-Harlig (1994) advocates a further exploration of these two particular cases, since they offer unambiguous evidence for either the Aspect or the Discourse Hypothesis – thus the Discourse Hypothesis would be confirmed if all foregrounded verbs, irrespective of their inherent lexical aspect, carried perfective markings in interlanguage; in contrast, the Aspect Hypothesis would be confirmed if all telic verbs, regardless of grounding carry perfective markings. Naturally, such an extreme scenario of the distribution of verb inflections could hardly ever be observed in a second language sample.

Grounding		
	Foreground	**Background**
Telic	AH = DH	AH ≠ DH
Atelic	AH ≠ DH	AH = DH

Table 3.6. Overlap between the Aspect Hypothesis (AH) and the Discourse Hypothesis (DH) (adapted from Bardovi-Harlig 1994: 55)

In sum, both the Aspect and the Discourse Hypothesis have established themselves as legitimate theoretical models for the analysis of universal developmental patterns in the L2 acquisition and use of tense-aspect morphology. Nevertheless, the Aspect Hypothesis is the more widely-used research framework of the two, and is still being used in various L2 samples, the most recent of which include computerised data comprising learner speech and writing – computer learner corpora (see chapter 4). Having dealt with the universal patterns in the development of L2 aspect morphology, the final section of this chapter will turn to a brief review of the possible forms of L1 influence on the L2 acquisition of aspect.

3.4. L1 Influence on the Acquisition of Aspect

The research frameworks outlined in the previous sections focus exclusively on universal developmental patterns of tense-aspect morphology in English as a second language, mostly irrespective of learner-internal factors such as the influence of the mother tongue or learner-external factors such as the effects of foreign-language instruction or the amount of L2 exposure. The present section focuses on the influence of L1 on the development of second-language temporality, reviewing studies that comment on its general role, as well as studies which investigate the specific role of the L1 tense-aspect system in the acquisition and use of L2 aspect. Slabakova claims that one important difference between the acquisition of aspect in L1 and L2 is the fact that "L2 acquirers already have a grammar in place and this grammar influences the L2 acquisition process in some way" (Slabakova 2000: 740), which goes beyond developmental, semantic or discourse constraints. With regard to the acquisition of temporality in L2 English by learners with different tense-aspect systems in their native languages, Comrie notes that "there is also a need to understand just how different the binding power of two typologically distinct languages may be […]. For instance, Spanish has a tense system rather similar to that in English, whereas Burmese is analyzed as having no tense system at all" (Comrie 1985: 307). Along similar lines, Rohde argues that "depending on the nature how tense and grammatical aspect are encoded in the languages involved, the acquisition of verbal inflections may support the Aspect Hypothesis to varying degrees" (Rohde 2002: 211).

The general influence of the native language on second language acquisition and use has long been the apple of discord in SLA research. Early contrastive studies categorize the L1 influence as "transfer" and claim that it occurs when "individuals […] transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture" (Lado 1957: 2). The majority of the early studies exploring the role of L1 in L2 acquisition and use follow the leading psychological school behind SLA research from the 1960s – Behaviourism (cf. Gass and Selinker 2001: 66; Odlin 1989: 15-17). Behaviourists identified two types of transfer depending on whether they resulted in correct or incorrect L2 use: 1) a positive type of transfer or transfer which facilitates learning, and 2) a negative type of transfer or transfer which interferes with L2 learning – the so-called interference (cf. Gass and Selinker 2001: 67). Negative transfer or

interference proved as the central concept behind behaviourist theories of SLA, since transferring different language habits from the mother tongue into the second language was held responsible for all learning difficulties and, ultimately, for all kinds of errors in the L2 output. These ideas served as the origin of the Contrastive Analysis Hypothesis (CAH), a powerful theoretical framework for comparing different languages and predicting learning difficulties and errors on the basis of the differences (cf. Lado 1957: 59).

A more recent theory on the role of L1 in L2 utilises a broader term – "cross-linguistic influence" – which is defined as "the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired" (Odlin 1989: 27). Odlin's transfer model relativises the role of the native language as the most important factor in second language acquisition and use and downplays the significance of the L1-L2 differences as the only source of errors in the L2 output. Furthermore, he argues that transfer is not necessarily synonymous with falling back on the native language, since knowledge of more than two languages may lead to "different kinds of source language influence, although pinning down the exact influences in multilingual situations is often hard" (Odlin 1989: 27). Therefore, he states that a fully-adequate definition of cross-linguistic influence should also include the concepts of learning processes and strategies, as well as simplification (cf. Odlin 1989: 28). In terms of the variable effects of cross-linguistic influence on the target language acquisition, Odlin admits that positive transfer is only determinable when the success rates of learning populations with different mother-tongue backgrounds are compared, whereas negative transfer is more easily determined by establishing divergences from the norms of the target language, mostly in terms of errors (cf. Odlin 1989: 36). These divergences can be subdivided into four different types:
(1) underproduction
(2) overproduction
(3) production errors
(4) misinterpretation

Underproduction refers to a situation where learners produce too few examples of a particular target-language structure, whereas overproduction describes the opposite phenomenon. Production errors can be further subdivided into three different subtypes: 1) substitutions, i.e. when learners use a native-language form in the target language, 2) calques, i.e. when learners literally translate a native-language structure into the second language, and 3) alterations of structures, i.e. when learners overreact to a particular native-language influence by trying to

hypercorrect a particular L2 form. Lastly, misinterpretations concern the erroneous perception of target-language sounds, word patterns or even cultural assumptions (cf. Odlin 1989: 37 – 38). The majority of the studies exploring the role of cross-linguistic influence focus on these four types of divergences, rather than comparing different groups of learners along the lines of positive transfer by looking at the similarities.

From a more narrow perspective, studies analysing the specific role of the L1 tense-aspect system on the acquisition and use of L2 aspect focus predominantly on the above-mentioned negative transfer effects, such as overproduction, underproduction and production errors. Several recent form-oriented studies (e.g. Collins 2002; 2004) explore the role of L1 transfer in relation to other constraints such as the inherent lexical aspect of verbs, and argue in favour of a "developmentally constrained L1 influence" (Collins 2004: 254). Collins (2002) found that beginning and intermediate French-speaking learners of English start overproducing the present perfect in inappropriate simple past contexts in a way that is formally similar to the use of the French *passé composé*, especially after they get comfortable with using the simple past (cf. Collins 2002: 254). However, she proposed that this particular L1 influence from French was only moderate and not the sole reason for incorrect uses of English verbal morphology and stated that it "cannot override the influence of inherent lexical aspect" (Collins 2002: 85). In a later study comparing L1 French and L1 Japanese learners of English (cf. Collins 2004), she discovered that both learner groups went through the same developmental stage of marking the simple past more frequently on telic verbs than on atelic verbs, and argued that L1 influence could only be detected in the case of French EFL learners' use of the present perfect with telic verbs (cf. Collins 2004: 267-268). Her findings support a "developmentally-constrained L1 influence" view where "the relative roles played by 'universal' semantic categories and 'particular' L1 knowledge" (Collins 2004: 252) can be best defined by comparing different learner populations acquiring the same target language. Along similar lines, Slabakova (2000) compared Bulgarian and Spanish low-proficiency EFL learners with regard to their awareness of the telicity of English verbs and found that Spanish low-proficiency learners were more successful in distinguishing telic from atelic verbs than Bulgarian low-proficiency learners. In contrast to Collins' "moderate view", she concluded that the differences between the two learner groups were "directly traceable to their native language" and the different telicity parameters in their respective L1s Spanish and Bulgarian (Slabakova 2000: 764). In a more recent cross-linguistic comparison between English and Japanese, Shirai and Nishi (2003) also emphasised the difficulty of acquiring inherent lexical aspect whenever there is a discrepancy between the learner's native language and the target

language. They suggested that both Japanese learners of English and English learners of Japanese will have problems acquiring verbs in the respective L2 in cases when there is no one-to-one correspondence between Japanese and English (cf. Shirai and Nishi 2003: 281). In a recent study on the acquisition of the imperfective aspect in L2 Japanese, Sugaya and Shirai (2007) also focused on the interplay between inherent lexical aspect and learners' L1 and found that L1 played a role in the distributional patterns predicted by the Aspect Hypothesis, but only to a certain extent, since it interacted with other variables such as the task type and learners' overall proficiency.

In sum, a number of the studies reviewed so far argue in favour of a moderate "cross-linguistic influence involving relativistic effects" (Odlin 2008: 306) – effects which range from developmental factors such as a universal reliance on verbal semantics (i.e. the Aspect Hypothesis) to learning-related variables such as the acquisitional environment and the proficiency levels of the learners. The next chapter will focus more specifically on the influence of the L1 tense-aspect system on the use of aspect in English as a foreign language – on the basis of learner corpus data.

3.5. Summary

The present chapter suggested that a number of factors affect the L2 development and use of aspect in English: on the one hand, the universal influence of the inherent lexical aspect of verbs and their distribution across different types of discourse; and on the other, learner-and learning-related variables such as the influence of the mother tongue and learners' proficiency levels. The state-of-the-art report presented in this chapter calls for a further investigation of the complex interplay between these factors, and especially with regard to the influence of the native language, which "has been curiously neglected so far" (Slabakova 2002: 184) in tense-aspect research altogether. As Shirai recommends, future tense-aspect research should address interlanguage comparisons focussing on "the differences between [e.g.] […] German learners (no progressive) vs. Chinese learners (restrictive progressive, with action in progress meaning only) in the acquisition of the highly grammaticised, polysemous English progressive […]" (Shirai 2009: 184). The goal of the present study is to fill this gap – by analysing the use of English aspect morphology by advanced learners of English with two radically different aspectual systems in their native languages. The next chapter will address some of the most recent form-oriented studies focussing on the use of aspect in learner language based on corpus data.

4. Corpus-Based Approaches to Aspect in Learner Language

This chapter will address the most recent formal approaches to the use of aspect in L2 English – by reviewing current research exploring aspect in L2 English on the basis of large computerised datasets consisting of samples of learner language – learner corpora. Corpus-based studies on the second-language acquisition and use of aspect analyse aspect as an observable phenomenon in learner speech or writing, both from a quantitative and a qualitative (form-oriented) perspective. The present chapter will set out with an overview of the benefits of corpus-based research for the study of grammar and its pedagogical implications for the foreign language classroom, and will proceed with a general description of learner corpora and their application in second language research. Lastly, the chapter will summarise the existing research findings on the use of aspect in L2 English on the basis of learner corpus data.

4.1. Corpora in the Study of Grammar

Although the main applications of corpora and corpus-based research have so far had the greatest impact on the writing of dictionaries (cf. Hewings and Hewings 2005: 82) and "grammar does not have a long tradition of empirical study" (Biber, Conrad and Reppen 1998: 55), corpus researchers (e.g. Meunier (2006: 37) and Conrad (2007: 55) argue that a number of corpus-linguistic developments have dramatically changed the study of grammar over the past 20 years. To illustrate, Meunier notes that "the benefits [of corpus-based research] for grammatical description are numerous" (Meunier 2007: 25) and identifies two major aspects of the corpus-based description of grammar:

(1) the identification of frequency
(2) the identification of grammatical patterns

Large-scale corpus-based analysis of language provides insights into the "frequency of grammatical or function words, parts-of-speech, grammatical phenomena and syntactic structures" (Meunier 2007: 25), whereas grammatical patterns can be highlighted through the recurrent lexico-grammatical combinations of words or more complex syntactic patterns (cf. ibid: 25). Biber, Conrad and Reppen (1998: 57-58) claim that corpus-based research on

grammar can be applied on all language levels and suggest four major research questions concerning the study of grammar which can be investigated with the help of corpora:

(1) How can the use and function of morphological characteristics [e.g. nominalisations] be better understood by analyzing their distribution across registers?
(2) How can the use and function of grammatical classes [e.g. the distribution of nouns and verbs] be better understood by analyzing their distribution across registers?
(3) How can the function of syntactic constructions [e.g. *to-* and *that*-clauses] be better understood by analyzing their distribution and linguistic associations across registers?
(4) What linguistic and non-linguistic features are associated with the choice between seemingly synonymous structural variants? [e.g. subject clauses vs. extraposed clauses – *That we mustn't worry too much is clear* vs. *It is clear that we mustn't worry too much*] (Biber, Conrad and Reppen 1998: 57- 58)

Answers to these four questions with respect to a wide range of grammatical phenomena have been provided in great detail in some of the most recent corpus-based grammars such as *LGSWE*, as outlined in section 2.2. (e.g. Biber et al. 1999). Corpus-based grammars place a special emphasis on the importance of different registers in English in contrast to the "grammar of English" as a monolithic concept (cf. Conrad 2007: 56), as well as on the explanation of linguistic functions with respect to the distribution of forms across different registers and varieties of English. Kennedy notes that some of the early corpus-based descriptions of English, "even before the age of computer" (Kennedy 1998: 122) already investigate the "various aspects of the distribution and use of verb-form morphology, prepositions, conjunctions and adverbials" (ibid.: 122) for purely pedagogical purposes – as "an indication of which [a]re the most frequently and widely used forms" (ibid.: 122) as a guide for both teachers and learners in the foreign language classroom. A comparison of three pre-computer-era research studies on the distribution of verb forms in different registers of English carried out manually on the basis of small corpora shows first tendencies concerning the distribution of finite forms in English and the eight most frequent tense-aspect uses in particular (see table 4.1).

Verb forms	Ota (1963)		George (1963a)		Joos (1964)
	Spoken US English (%)	Written US English (%)	UK English plays (%)	Written UK English (%)	Written UK English
simple present	64.4	26.4	67.6	38.4	39.6
simple past	18.3	58.5	14.4	48.2	31.3
present perfect	4.8	2.7	5.3	3.1	4
past perfect	0.4	3.4	0.9	4.1	2
present progressive	5.4	0.9	4.4	1.4	2.2
past progressive	0.9	1.1	0.4	1.4	2.2
present perfect progressive	0.5	0.1	0.6	0.1	0.1
past perfect progressive	0.01	0.2	-	0.1	0.1
other verb forms	5.3	6.6.	6.4	3.2	18.4

Table 4.1. Relative frequencies of use of finite verb forms (adapted from Kennedy 1998: 126)

To illustrate, the comparison shows that the simple present tense dominates spoken language, whereas the simple past dominates written language; further, both the progressive and the perfect are much less frequent than the simple aspect forms, the perfect being in general more frequent than the progressive and the present perfect and the present progressive being more frequent than their past counterparts. Francis and Kučera (1982: 555) confirm these tendencies with respect to the distribution of the progressive and the perfect by identifying their relative frequencies in the first one-million-word corpus of American English featuring a collection of different written registers – the *Brown corpus* (cf. Francis and Kučera 1979).

Genre	Perfect forms		Progressive forms	
	Number	%	Number	%
A. Press: reportage	469	5.94	297	3.76
B. Press: editorial	367	6.99	231	4.40
C. Press: reviews	188	5.96	88	2.79
D. Religion	210	6.16	73	2.14
E. Skills and hobbies	326	4.92	142	2.14
F. Popular lore	588	6.32	251	2.70
G. Belles lettres	1,075	7.15	362	2.41
H. Miscellaneous	233	5.45	90	2.10
J. Learned	739	5.54	210	1.57
K. General fiction	563	7.45	320	4.24
L. Mystery and detective	511	7.52	268	3.94
M. Science fiction	109	6.78	52	3.24
N. Adventure and western	518	6.40	328	4.06

P. Romance and love story	601	7.30	348	4.23
R. Humour	148	6.82	83	3.82
whole corpus	**6,645**	**6.47**	**3,143**	**3.06**

Table 4.2. Perfect and progressive verb forms in the Brown Corpus (adapted from Francis and Kučera 1982: 555 in Kennedy 1998: 127)

"Calculating the relative frequency of occurrence of a particular grammatical feature is, of course, only the starting point of cross-corpus comparison" (Hewings and Hewings 2005: 84) and the distribution results of these early corpus-based studies on grammar have been overhauled by more comprehensive reports addressing all four of Biber et al.'s (1998) research questions, e.g. reporting on the linguistic association between certain tense-aspect forms and certain lexical verbs, establishing the most 'present-tense prone' lexical verbs (cf. Kjellmer 1992 in Kennedy 1998: 128), identifying the most frequent functions of tense-aspect forms and analysing them in context etc. (e.g. Biber et al. 1999; Mindt 2000).

The corpus-based findings on grammatical patterns in English have fed into the creation of new teaching materials for the EFL classroom, affecting textbooks and reference grammars, as well as individual teachers' decisions on what to teach, how to teach and when to teach it (cf. Römer 2008: 112 – 113). Regarding the latest developments of corpus-based grammar research and its applications to teaching, Conrad (2007: 56) argues that "three changes prompted by corpus-based studies of grammar have the potential to revolutionize the teaching of grammar":

(1) Monolithic descriptions of English grammar will be replaced by register-specific descriptions
(2) The teaching of grammar will become more integrated with the teaching of vocabulary
(3) Emphasis will shift from structural accuracy to the appropriate conditions of use for alternative grammatical constructions (Conrad 2007: 56)

While register-based and even variety-based grammatical descriptions of English have already fed into corpus-based reference tools like *LGSWE* (cf. Biber et al. 1999), there are virtually no pedagogical or EFL grammars which derive their findings from non-native speaker data. Meunier (2007: 26) recommends an approach where the combined evidence from authentic native as well as non-native English data should be used for the "curriculum design, the production of reference tools and classroom EFL grammar teaching" and notes that the insights into the problematic areas of English grammar for EFL learners should be the basis for a "more focused and appropriate teaching" (ibid.: 27). The benefits of applying

corpus techniques to non-native data in order to highlight learners' problems in particular areas of English grammar, as well as their pedagogical implications will be dealt with in detail in the next section.

4.2. Learner Corpora

The new advances in corpus-based research have revolutionised not only the study of grammar, but also all fields of "language-related research, from lexicography to literary criticism through artificial intelligence and language teaching" (Granger 1998: 3). A great number of SLA researchers have called for larger and more systematic datasets than the small elicited and introspective learner samples used in traditional second-language acquisition research (cf. Ellis and Barkhuizen 2005: 23-38). The purpose of such larger samples is that researchers can better prove whether "the results obtained are applicable only to the one or two learners studied, or whether they are indeed characteristic of a wide range of subjects" (Gass and Selinker 2001: 31). Corpus linguistics and the development of corpora have provided both for a new source of language data and for a research methodology which allows for a direct look at the learners' output in order to answer the following questions:

(1) How does interlanguage pattern?
(2) Which (kinds of) errors do the language learners commit […]? (cf. Gries 2008: 413)

This direct look has been made possible by the invention of learner corpora – computerised electronic collections of foreign or second language learner texts assembled in order to "investigate learner language in a way that was not possible previously" (Pravec 2002: 81). Granger (2002: 7) proposes the following definition of computer learner corpora (CLC), which will be used henceforth:

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.

Applied to learner data, the underlying notion of authentic language usage is somewhat different to the notion used for native-speaker performance data since it covers "different degrees of authenticity" (Granger 2002: 8), where the various ways of student-teacher communication in the EFL classroom (e.g. essay writing or reading foreign language texts aloud) qualify as authentic learner data (cf. Granger 2002: 8). Granger's definition treats the

term "foreign language" (FL) as a term relating to "the learning of a non-native language in the environment of one's native language" (Gilquin and Granger 2011: 56), whereas the term "second language" (SL) is reserved for the learning of a non-native language in the environment of this language, e.g. English in Britain (cf. Granger 2002: 8; Gilquin and Granger 2011: 56). The explicit design criteria refer to the variables specific to a particular learner corpus, which can be grouped into two major groups:

- learner-related variables (e.g. mother tongue, level of proficiency etc.)
- task-related variables (e.g. exam situation etc., summarised in table 4.3).

Lastly, the particular SLA/FLT purpose is to prove or disprove specific aspects of SLA theory on the basis of these authentic learner corpus data (cf. Granger 2002: 9-10).

LEARNER	TASK SETTINGS
• Learning context	• Time limit
• Mother tongue (L1)	• Use of reference tools
• Other foreign languages	• Exam
• Level of proficiency	• Audience/interlocutor
[…]	[…]

Table 4.3. CLC specific design criteria (adapted from Granger 2002: 9)

Learner corpora can be standardised and annotated in a different manner, ranging from plain text corpora to annotated corpora enriched with textual meta-information such as grammatical categories (e.g. parts-of-speech tagging) or syntactic structure (parsing) (cf. Granger 2002: 10; Hunston 2002: 22-24)[21]. Granger (2004: 292) differentiates between two major types of learner corpora – commercial learner corpora and academic learner corpora, which can be further subdivided into multilingual (Multi-L1) learner corpora containing data from a wide range of EFL/ESL learners with various mother-tongue backgrounds, and monolingual (Mono-L1) learner corpora containing data from just one learner population. Commercial learner corpora like the Longman Learners' Corpus and the Cambridge Learner Corpus have been widely used to inform EFL materials like dictionaries, textbooks and grammars for the so-called delayed pedagogic use (e.g. describing the general learner problems found in the learner output by many learner populations and targeting them in the subsequently published reference tools). Academic learner corpora can be used for both immediate and delayed pedagogical purposes – the first case is identical to the delayed pedagogical use of commercial learner corpora, whereas in the second case learners can

---

[21] The majority of the learner corpora available so far are in plain text format; the subsequent types of annotation used for the corpus data in the present study will be outlined in more detail in chapter 5.

produce and study the learner corpus data at the same time (cf. Granger 2004: 292 – 293). The *International Corpus of Learner English*, which is used as a database for the analysis of the present study is one of the few academic learner corpora that are multilingual and that can be used for both immediate and delayed pedagogical purposes in the EFL classroom (e.g. for the "improve your writing skills" sections in the second edition of the Macmillan English Dictionary, cf. De Cock et al. 2007: IW1-IW50).
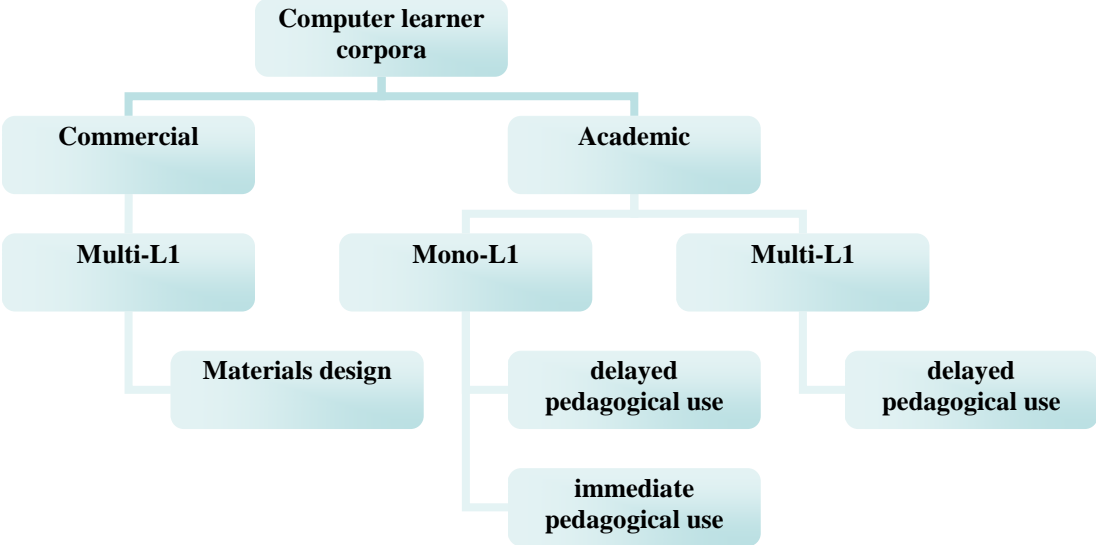
Figure 4.1. Computer learner corpora (adapted from Granger 2004: 292)

Learner corpus research belongs to the newest developments of corpus-based research which uses learner corpora as a starting point of investigation and helps to establish a link between four major fields in language research: Corpus Linguistics, Linguistic Theory, Second Language Acquisition and Foreign Language Teaching (cf. Granger 2009: 15).
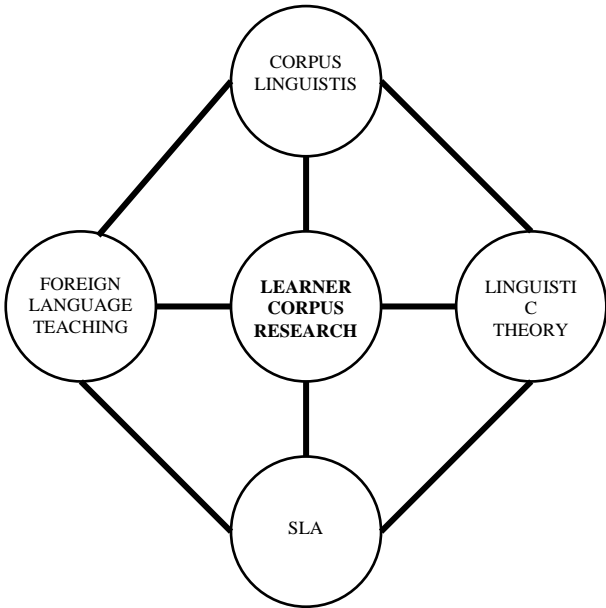
Figure 4.2. Core components of learner corpus research (adapted from Granger 2009: 15)

Thorough knowledge of these four fields is necessary in order to "provide the […] underpinnings for linguistic analysis of the data" (Granger 2009: 15), where corpus linguistic expertise combined with the knowledge of linguistic theory and second-language acquisition research is the stepping stone to interpreting the results and leads to "effective pedagogical applications" (Granger 2009: 16) in the foreign language classroom. Learner corpora bring about a number of research questions as a starting point for qualitative and quantitative investigations of learner corpus data, as outlined by Leech (1998: xiv):

(1) What linguistic features in the target language do the learners in question use significantly more often ('overuse') or less often ('underuse') than native speakers do?
(2) How far is the target language behaviour of the learners influenced by their native language (NL transfer)?
(3) In which areas do they tend to use 'avoidance strategies', failing to exploit the full range of the target language's expressive possibilities?
(4) In which areas do they appear to achieve native-like or non-native-like performance?
(5) What (in order of frequency) are the chief areas of non-native-like linguistic performance which learners in country A suffer from and need particular help with?

The research method pursuing the answers of the above questions falls under the Contrastive Interlanguage Analysis (CIA) framework developed by Granger (1996), which has been developed as a new type of contrastive analysis with the aim to compare learner and native varieties of the same language or "what non-native and native speakers of a language do in a comparable situation" (Pery-Woodley (1990:143) in Granger 1996: 43). Contrastive Interlanguage Analysis does not compare original data from one language with original data from another language like traditional Contrastive Analysis (cf. Gilquin 2008: 6), but rather the different non-native varieties of the same target language (e.g. English). The CIA approach is at the heart of the *ICLE* corpus since its ultimate goal is "to uncover factors of 'foreign-soundingness' in learner writing" (Granger 1996: 43). CIA involves two types of comparison and is diagrammatically represented in figure 4.3 (cf. Granger 1996: 44):

(1) Native language (NL) vs. interlanguage (IL) – a comparison between native and non-native varieties of one and the same language, e.g. native and learner English
(2) Interlanguage (IL) vs. interlanguage (IL) – a comparison between different interlanguages of the same language, e.g. the English or French learners vs. the English of German learners

Figure 4.3. Contrastive Interlanguage Analysis (adapted from Granger 1996: 44)

The two types of comparison within the CIA framework are only possible with comparable learner and native data, where variables such as text type and learner type are strictly controlled and well documented – a CIA of biased data comprising different text types and different learner proficiency levels would yield unreliable results. While the second comparison has remained largely undisputed, the first type of comparison has been the object of a heated argument due to its failure to analyse interlanguage as a language system in its own right and is known as the "comparative fallacy" (Bley-Vroman 1983 in Granger 2009: 18). Granger (2009: 18-19) argues against the comparative fallacy (calling it "comparative hypocrisy") and points out that all traditional SLA studies analyse learner language with an implicit notion of a native-speaker target norm, emphasising that the CIA approach can still highlight the features of learner language "from a strictly L2 perspective" (Granger 2009: 18).

The quantifiability of learner and native corpus data is what makes CIA particularly attractive as a methodological framework. Learner and native corpus data make it possible for the researcher to carry out quantitative contrastive comparisons on the basis of selected linguistic features, by identifying instances of learner overuse (deviations in plus from the corpus-based native norm) and learner underuse (deviations in minus from the corpus-based native norm) (cf. Granger 1996: 45). In order to predict which features might deviate in plus or minus from the native-speaker target norm, traditional Contrastive Analysis is used to compare the respective features in the original languages before coming to the learner varieties of the same target language – a model known as the Integrated Contrastive Model (cf. Granger 1996: 47; Gilquin 2008: 8). The Integrated Contrastive Model (represented diagrammatically in figure 4.4) is a unique model since it combines the predictive potential of traditional Contrastive Analysis with the diagnostic potential of Contrastive Interlanguage Analysis. The prediction is carried out by comparing a certain feature in two original source languages (e.g. the frequency of the passive in French and in English) or in a source language SL (e.g. French) with a translated language TL (e.g. English), the assumption being that differences in the frequencies of use (e.g. more passives in English than in French) are likely to result in a negative transfer (e.g. the passive is likely to be underused by French EFL learners). The diagnosis is given by comparing the learner data with the native data, as well as comparing different learner datasets (e.g. produced by learners with different L1s – e.g. French and Spanish EFL learners) with each other in order to establish L1-specific deviations (e.g. compare whether French learners' underuse of the passive is more significant than Spanish learners' underuse of the passive) (cf. Granger 1996: 46 – 47; Gilquin 2008: 7 – 8).
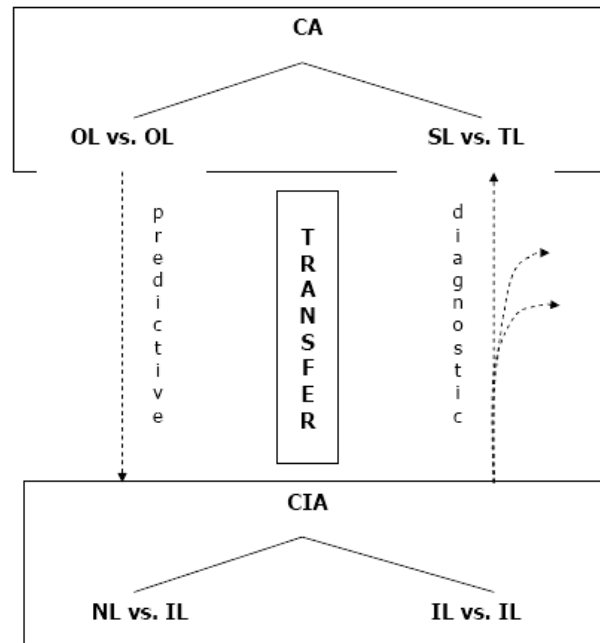
Figure 4.4. Integrated Contrastive Model (Granger 1996: 47)

The Integrated Contrastive Model as a combination of two underlying models is best suited for the identification of transfer in the foreign language, since "[t]ransfer is a slippery phenomenon that does not lend itself easily to apprehension" (Gilquin 2008: 25) and needs contrastive evidence from different datasets. In addition to the instances of negative transfer in the form of overuse and underuse identified with the help of this model, learner corpora contain many instances of misuse or "anomalous" learner use (cf. Granger 2002: 18). One way to detect and quantify the instances of learner misuse is to annotate the learner corpus data for errors – for example by using the Louvain error-tagging system (cf. Dagneaux et al. 1998; Dagneaux et al. 2008). The Louvain error-tagging system includes an error-tagging manual which illustrates the hierarchy of error categories and subcategories (e.g. grammatical errors are marked with a G, whereas grammatical errors affecting verbs are marked with GV) and an error-tagging computer tool – the Louvain Error Editor (Dagneaux et al. 2008), which allows for an upload of the learner corpus data and an automatic insertion of the respective error tags (cf. Granger 2002: 19). The Louvain error tagging procedure is based on initial error detection and correction carried out by a native informant and a subsequent error tag insertion carried out by a linguist (cf. Dagneaux et al. 1998: 165). The advantages of annotating learner corpus data for errors are enormous, even though rarely exploited in the existing research so far: the annotated data can be used for the computer-aided error analysis

approach proposed by Dagneaux et al. (1998), where learner misuse can be analysed alongside correct learner use[22].

Identification of overused, underused or misused patterns is not the last step in learner corpus research – Leech (1998: xix –xx) notes that not all of these deviations "should feed into the development of teaching materials" as a final pedagogical application of the results. Along similar lines, Granger argues for a flexible approach towards incorporating those findings into the curriculum which are suitable for the learning goals of the learners; thus, overuse, underuse and misuse should be dealt with in greater detail only in cases when learners want "to achieve near-native proficiency but can clearly be neglected or simply presented as useful strategies for learners whose language learning aims are less ambitious" (Granger 2009: 22).
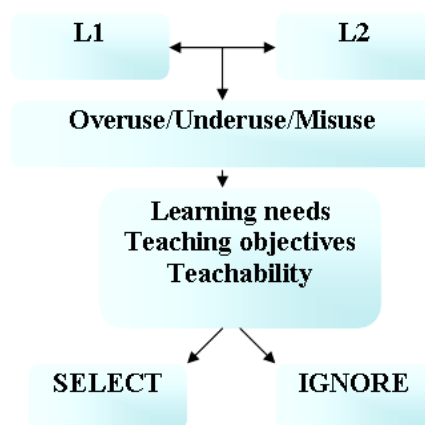
Figure 4.5. From learner corpus analysis to language teaching (Granger 2009: 23)

To summarise, learner corpus studies may have numerous (delayed) pedagogical applications: they can be used to inform teaching materials such as EFL dictionaries, textbooks and grammars (cf. Granger 2004; Römer 2008); in addition, they can be used for immediate pedagogical purposes as hands-on data within a data-driven-learning approach in the EFL classroom[23]. In addition, they can be used for syllabus design (cf. Aston 2000) insofar as they highlight the needs of a particular learner population (cf. Meunier 2002: 125) and reveal whether or not a particular language feature is difficult for a particular learner population and whether it should be specifically targeted in the EFL classroom (cf. Granger 2002: 22).

---

[22] Although minimised, the "fallacies" of traditional Error Analysis still hold for such a computer-aided error-tagging approach, since native informants involved in the tagging still have implicit target language norms, which may vary from (native) speaker to speaker, see also chapter 8

[23] The delayed pedagogical applications of learner corpora outweigh by far the immediate pedagogical applications so far – Mukherjee and Rohrbach (2006: 206) note that "there is a widening gap and a growing lag between on-going and intensive corpus-linguistic research on the one hand and classroom teaching on the other"

Nesselhauf (2007: 305) argues that "the path from learner corpus analysis to language pedagogy is not [as] direct as sometimes appears to be assumed" and calls for a greater focus on the frequency, difficulty and degree of deviation of linguistic items in learner language in the EFL courses targeting the advanced learner of English (cf. Nesselhauf 2007: 311). The next section will focus specifically on the existing learner corpus research on the above-mentioned deviations (overuse, underuse and misuse) with regard to the use of aspect forms by EFL learners.

4.3. Learner Corpus Research on the Progressive and the Perfect

Alongside the form-oriented studies outlined in chapter 3 which focus on the acquisition of aspect in L2 and identify universal patterns in its development, irrespective of the mother-tongue backgrounds and proficiency levels, several corpus-based studies investigate specifically the use of the progressive and the perfect on the basis of learner corpora, mostly following the Contrastive Interlanguage Analysis framework outlined in the previous section. The present section will briefly summarise these findings.

Hinkel (cf. Hinkel 2004: 5) notes that even advanced EFL learners after many years of learning and constant use still have problems with the conventionalized uses of tenses and aspect, especially in academic settings and in particular in writing. Most learner corpus studies so far have focused on the use of the progressive in L2 academic writing (with the exception of Davydova (2011) who analysed non-native uses of the perfect); in addition, some of the studies examine the general distribution and use of tense-aspect forms in learner writing. However, none of the learner corpus studies published so far has focussed on the use of aspect in particular (most of them analyse learner use of aspect forms as part of a broader analysis involving a whole range of tense-aspect forms), and especially on its use by different learner populations with different aspect systems, but at the same (advanced) level.

One of the early learner corpus studies to examine the general patterning of tense-aspect forms in learner language is Granger's (1999) study on an error-tagged pilot subcorpus of the *International Corpus of Learner English* (*ICLE*)[24] (cf. Granger et al. 2002) consisting of learner essays written by both upper-intermediate and advanced French-speaking EFL learners. Granger found out that a large number of the tense-aspect errors in learner writing

---

[24] for a detailed description of the International Corpus of Learner English, see chapter 5

were due to learners' clause-level approach to tense and aspect and their subsequent inability to maintain tense continuity beyond clause boundaries or in a longer stretch of writing. In addition, she observed that certain tense-aspect combinations such as the past progressive or the present perfect were more prone to errors than others and developed very little from the post-intermediate level to the advanced level. Granger attributed the particularly high rates of misuse of these forms to possible transfer from learners' native L1 French, commenting on the contrasts between the tense-aspect systems of the two languages (e.g. the dynamicity of English vs. French stativity, cf. Granger 1999: 199) and pointed out that tenses must be taught contrastively as part of a tailor-made teaching approach, as well as from a discourse, rather than a sentence-based perspective (cf. Granger 1999: 198-202). Along similar lines, Abe and Tono (2005) examined the use of tense and aspect in two error-tagged learner corpora of written and spoken learner English produced by Japanese learners of English. They identified a strong correlation between written production and the high proportion of tense-aspect errors and concluded that Japanese learners experienced considerable difficulties preserving a coherent sequence of tenses in written texts. Similar to French-speaking learners of English, Japanese learners misused the morphologically more complex tense-aspect forms such as the present perfect, the past perfect or the present progressive more frequently than the simpler forms like the simple present or the simple past (cf. Abe and Tono 2005: 9-10).

A number of learner corpus studies focus specifically on the L2 use of the progressive aspect, which they acknowledge as a major "challenge for language learners, even for advanced levels, and particularly for learners whose L1 does not have a direct counterpart to the progressive (Wulff and Römer 2009: 116). Axelsson and Hahn (2001: 5) note that:

> [t]he progressive is a feature of English grammar that is difficult to handle for non-native speakers, both teachers and students. One consequence is that the progressive is claimed to be used too often and in the wrong places by Swedes and Norwegians.

Learner uses of the progressive have been investigated so far mostly on the basis of learner corpora comprising written production of learner populations with Germanic mother-tongue backgrounds. Virtanen's study (1997) is one of the few (early) exceptions which compare the use of the progressive by learners with both Germanic and Non-Germanic L1 backgrounds – in two subcorpora of the Finnish and Swedish components of *ICLE*. Virtanen found significant differences in the frequencies of use of the progressive between these two learner corpora and also in comparison with two comparable native-speaker corpora of American and British English. She acknowledged that the differences were more significant

in the case of the comparison between the learner data and the American data and proposed that learner use of the progressive be compared with the corpus-based usage norms in both British and American English (cf. Virtanen 1997: 308 – 309). Axelsson and Hahn (2001) investigated the progressive in the Swedish and German components of *ICLE* and attested no significant differences between Swedish and German learners' use, in contrast to the significant differences found between the frequencies of occurrence of the progressive in British and American expert native writing. Although both German and Swedish as native languages lack the progressive as a fully-fledged grammatical category and were thus expected to cause underuse in Swedish and German learner English, such underuse did not occur: both learner groups used the progressive either within the native frequency span or slightly overused it (cf. Axelsson and Hahn 2001: 11-12). Still, a detailed qualitative analysis of the German and Swedish learners' use of the progressive showed that there were many instances of "ambiguous uses" of the progressive (Axelsson and Hahn 2001: 17) – uses which were judged by native speakers of English as not entirely incorrect, but as marginally acceptable within the discourse context of a given learner essay. Similar to Virtanen's (1997) and Axelsson and Hahn's (2001) findings, Lenko-Szymanska (2007) found a considerable overuse of the past progressive in the Polish component of *ICLE*; in contrast to Axelsson and Hahn (2001), she attributed this overuse, and in particular the distributional bias of the past progressive over simple past forms in the Polish learner corpus to interference from Polish. Lenko-Szymanska suggested that this overuse was due to the fact that Polish learners associated the Polish imperfective past with the English past progressive, overusing the latter to relate the background of an event or a situation as they would have done with the imperfective past in their native Polish (cf. Lenko-Szymanska 2007: 264). In terms of the correlation between the past progressive and the inherent lexical aspect of verbs, Lenko-Szymanska found no significant differences between the learner and the native-speaker distribution – both Polish learners and native-speakers showed a certain tendency towards using more past progressives with activity verbs, although to a slightly different extent (cf. Lenko-Szymanska 2007: 262) – thus confirming the claims of the Distributional Bias Hypothesis proposed by Andersen and Shirai (1996)[25].

This reported overuse of the progressive in non-native English has also been examined by a couple of more recent corpus-based studies which compare learner varieties of English as

---

[25] Wible and Huang (2003) attest a similar distributional bias with regard to the overuse of simple past markings with telic verbs in both a native-speaker corpus and a learner corpus of advanced Taiwanese learners' written English

a foreign language (EFL) with second-language varieties of English (ESL). Hundt and Vogel (2011) investigate the use of the progressive in German, Swedish and Finnish learner English with several ESL varieties of English (e.g. Singapore, Malaysian and Kenyan English, among others) and several native (ENL) varieties of English (e.g. British English and New Zealand English) in order to identify common patterns between these different varieties along an estimated ENL – ESL – EFL cline (cf. Hundt and Vogel 2011: 146). Contrary to their expectations, they acknowledge that all three varieties – ENL, ESL and EFL do not group together, but rather display a considerable internal variance, with New Zealand English being the leading native variety "with clearly the most frequent use of progressives in both unpublished and student academic writing" (Hundt and Vogel 2011: 155), German learner English the leading EFL variety with the most progressives, and Fijian English the ESL leading variety (cf. Hundt and Vogel 2011: 154-155). Moreover, Hundt and Vogel establish a "stretched tolerance" (Hundt and Vogel 2011: 158) for some ESL varieties in terms of their preference for the progressive with stative verbs, a finding which contradicts the fourth claim of the Aspect Hypothesis (the progressive is not incorrectly overextended to stative verbs) and which they reject for the learner Englishes and some of the native English varieties in their sample.

The increased frequency of the progressive attested by the corpus-based studies mentioned above is thus often accompanied by a functional shift, which has also been discussed as an "extended use" of the progressive (e.g. Comrie 1976: 38; Römer 2005). Römer (2005) found a similar functional extension of the progressive in corpora of spoken (native) British English, which she classified into seven additional functions, including 1) general validity, 2) emphasis and attitude and 3) gradual change and development (cf. Römer 2005: 95). In the contexts of English as a non-native variety, the extended progressive has also been identified as "a kind of continuous aspect without temporal immediacy" (van Rooy 2006: 37) and as an "attractive progressive" (Ranta 2006: 112), a form which carries a higher "communicative value in interaction" (Ranta 2006: 112) and which is easily recognised by learners and ESL users as such. Ranta argues that the extended progressive is a typical feature of English as a lingua franca (as spoken by speakers of many different languages), mainly because of its salience and "attention-catching form" (Ranta 2006: 114). She observes that non-targetlike uses of the progressive are not problematic because they do not lead to communication breakdown (cf. Ranta 2006: 114). However, even if research on the development of the progressive in British and American English has shown that the progressive has become more frequent, gradually expanding its meaning over the past two

centuries and especially in the past forty years (e.g. Hundt 2009; Leech and Smith 2006; Mair and Hundt 1995; Smitterberg 2005), some of the extended uses of the progressive found in the language of EFL learners are still treated as problematic in SLA research and remain an object of investigation in both traditional SLA research and in learner corpus studies.

Two of the most recent corpus-based studies examine learner uses of the progressive from a slightly different perspective, focussing on the native language input and its individual features such as the frequency effects and the lexico-semantic associations between verbs and the progressive, claiming that these features influence the L2 use of the progressive (cf. Wulff et al. 2009; Wulff and Römer 2009). Wulff et al. (2009) suggest that verbs that occur frequently in the progressive and that are more strongly associated with it in the native-language input are also acquired and produced earlier in the progressive by learners of English (cf. Wulff et al. 2009: 365). Wulff and Römer (2009) identify different degrees of association between lexical verbs and the progressive in learner writing and in different types of native-speaker writing, by comparing learner corpora with corpora comprising native novice and native expert writing. They acknowledge that German EFL learners show a greater preference for activity and motion verbs like *run* and *walk* in the progressive, as well as tend to overgeneralise the progressive and extend it to stative verbs like *miss* and *lack* to a greater extent than native speakers (cf. Wulff and Römer 2009: 121). However, they conclude that depending on academic expertise and learner proficiency level, both learners and native writers move along a semantic continuum with regard to their use of the progressive, which ranges from physical activities (preferred by learners and inexperienced writers) to communication verbs. This semantic continuum is accompanied by a functional shift from more core meanings to more modal meanings, and a grammatical shift from a more narrative past–time reference to a more impersonal present tense usage (cf. Wulff and Römer 2009: 130).

Other recent corpus-based studies address the progressive (and the perfect, albeit to a lesser extent) from a more general perspective, investigating it as one out of many tense-aspect forms in the development of the L2 verb system (e.g. Housen 2000; Housen 2002a; Housen 2002b). Housen (2002b) used both longitudinal and cross-sectional corpus data to compare the progress French and Dutch-speaking learners of English made in three years in terms of their development of tense-aspect morphology and found that the use of the *–ing* form gradually decreased with increasing proficiency of the learners in his sample. Further, he observed another counterexample of the fourth claim of the Aspect Hypothesis in the case of

an initial overuse of the *–ing* marker with stative verbs of cognition such as *know* or *seem* (cf. Housen 2002a: 104). In addition, both his native-speaker control corpus and the French and Dutch EFL learners' corpora showed a similar distributional bias with respect to the tendency of activity verbs to receive progressive markings – a finding which he explained with the fact that learners analysed "the *–ing* morpheme in terms of universal *prototypical* meanings first" (Housen 2002a: 108, original emphasis), thus confirming the prototype account for the acquisition of the progressive (Andersen and Shirai 1995). Similar to Wulff et al. (2009) and Wulff and Römer's (2009) findings, Housen suggested that the interplay of three major factors was responsible for the tense-aspect development in L2 English: transfer from the native language, the principles of language processing, as well as the frequency and distribution of verb forms in the input language (cf. Housen 2002a: 107-108).

In contrast to the (comparative) multitude of studies dealing with the progressive, the only comprehensive corpus-based study which focuses exclusively on the L2 use of the perfect published so far is Davydova's (2011) study. Davydova compares the use of the present perfect in traditional L2 varieties of English such as Indian English and Singapore English with learner varieties of English such as Russian learners' English and German learners' English, trying to find similarities between these different varieties and thus unify the contrasts between Outer Circle and Expanding Circle Varieties of English. She employs a multivariate analysis of the Outer and Expanding Circle corpus data in order to account for a multitude of factors, ranging from extra-linguistic variables such as gender of the speakers to language-internal factors such as transitivity, the use of time adverbials and the inherent lexical aspect of verbs. Davydova's findings show that both Russian and German EFL learners use the present perfect sparingly, applying it in approximately 30 % of all present perfect contexts; moreover, both learner groups in her sample seem to be strongly influenced by the inherent lexical aspect of verbs to the extent that they prefer only dynamic verbs in the present perfect (cf. Davydova 2011: 289). At the same time, she identifies differences between the two learner populations: to illustrate, German EFL learners seem to prefer "past tense morphology as a default strategy in order to avoid the ambiguous HAVE-perfect" (Davydova 2011: 289), whereas Russian EFL learners experience problems with the semantic complexity of the present perfect and deviate strongly from Standard English native speakers' use with respect to adverbial specification (cf. Davydova 2011: 272). Davydova's conclusion is that transfer from the native language seems to be the reason behind learners' non-targetlike use of the present perfect in both cases, although the different native language backgrounds

(L1 Russian and L1 German) result in a similar outcome of avoidance of the present perfect and its replacement by the simpler form – the simple past (cf. Davydova 2011: 290).

Another recent comprehensive study on learners' use of tense and aspect so far is Eriksson's (2008) corpus-based study which focuses on the use of tense, aspect and their combination in advanced Swedish EFL learners' written English. Eriksson describes the use of L2 verbal morphology on the basis of the Swedish component of *ICLE*, analysing advanced Swedish EFL learners' use of tense-aspect forms, both from a quantitative and a discourse perspective. In quantitative terms, Eriksson identified significant differences between the learner and the native-speaker frequencies of use with regard to the distribution of finite verb forms in the learner and native corpora – a finding which he attributed to the considerable variation between the essay topics in the learner and control corpora on the one hand, and to Swedish learners' "verbier" style of writing on the other (cf. Eriksson 2008: 219 – 220). In addition, he found numerous instances of learner misuse of the widely-quoted problematic tense-aspect areas such as the present progressive and the present and past perfect; notably, he observed that especially the progressive accounted for the majority of instances of aspect misuse in advanced Swedish learners' English. Similar to the findings of the studies mentioned above, he identified cases where the progressive was overgeneralised to stative and habitual situations and used with inanimate subjects to express states or habits of general validity, thus disproving once again the fourth claim of the Aspect Hypothesis (cf. Eriksson 2008: 221-222). In line with Granger's (1999) findings, he reported frequent unmotivated tense shifts in the Swedish learner corpus, which he attributed to learners' sentence-level approach to tense and aspect, and also called for a stronger discourse-based teaching approach to tense and aspect in the Swedish EFL classroom. Eriksson's conclusion is that Swedish EFL learners' misuse of tense and aspect can be explained by a combination of factors such as "transfer, overgeneralisation and inadequate discourse strategic skills" (Eriksson 2008: 222).

One last very recent study which deals with learners' misuse of tense-aspect forms in particular is Meunier and Littre's (2013) study based on the combination of learner-corpus data with experimental data. Meunier and Littre analysed the development of tense-aspect errors in the output of advanced French-speaking EFL learners over a period of 3 years, combining longitudinal learner corpus data with experimental grammaticality-judgement data in order to better observe, explain and gain a deeper understanding of advanced learners' error patterns and the various factors that come into play in the late stages of the L2 verb system

development. They found that while French EFL learners' errors decreased over time, certain features remained "unmastered" even at an advanced level: more than 50% of the errors they found were "aspect-only errors" (cf. Meunier and Littre 2013: 68), followed by tense-only errors (25%) and combined tense-aspect errors (20%). Based on the analysis of the accuracy of learners' use of the progressive, Meunier and Littre concluded that advanced EFL learners had obvious problems grasping less-salient features of the progressive than its core meaning "ongoing extendedness", like e.g. the use of the progressive for future (e.g. planned) actions (cf. Meunier and Littre 2013: 72).

4.4. Summary

To summarise, the learner corpus studies on the use of the progressive and the perfect in L2 English reviewed thus far reveal that, in general, EFL learners' use of English aspect forms clearly deviates from the native-speaker corpus-based norm: notably, learners from a variety of mother-tongue backgrounds such as Dutch, Finnish, French, German, Polish, Russian and Swedish all encounter considerable difficulties, even at an advanced level, and overuse, underuse or misuse the English progressive and perfect aspect in written (and other forms of) L2 English.

Notably, target-like use of the progressive and the perfect in L2 English is not achieved even at a very advanced level and after many years of instruction and exposure to English: to illustrate, the majority of the learner corpus studies (as well as a number of the more recent form-oriented studies on the acquisition of aspect reviewed in section 3.3) so far identify numerous instances of non-targetlike use of aspect forms at an advanced level of learning, which they attribute to a multitude of factors such as transfer effects, lack of writing skills and register awareness, as well as overgeneralisation and simplification strategies. Two general tendencies concerning the frequencies of use of the progressive and the perfect in advanced EFL learner writing emerge here:

(1) the tendency for advanced EFL learners to overuse the progressive
(2) the tendency for advanced EFL learners to underuse or avoid the present perfect and replace it by other tense-aspect forms, such as e.g. the simple past

The reasons behind these two tendencies appear to be likewise manifold: they range from the general attractiveness and salience of the progressive for EFL learners (i.e. 'attractive,

typically English progressive') to the semantic ambiguity of the English present perfect (and its comparatively late acquisition in L2 English). The present study aims to examine these two tendencies, by investigating the use of the progressive and the perfect contrastively in the writing of two learner populations at an advanced level, with two radically different mother-tongue backgrounds – Bulgarian and German. In addition to the quantitative corpus-based analysis and comparison of use of aspect forms in advanced Bulgarian and German EFL writing, the present study will carry out a contrastive qualitative analysis of the types of learner misuse of aspect forms, in an attempt to unify traditional corpus linguistic methods with a more qualitative SLA approach. To this end, the present study will draw on the findings quoted so far, and in particular with regard to the German learner data[26]; at the same time, it will be the first learner corpus study of its kind to carry out a systematic quantitative and qualitative comparison between aspect use in two learner corpora featuring learner populations with different aspectual systems in their native languages that have not been compared thus far and that still remain a research desideratum (cf. Shirai 2009: 184).

---

[26] There are no comparable studies focussing on the use of aspect in Bulgarian learner writing so far; therefore, the present study is the first of its kind to focus on Bulgarian learner data (see also the next chapter)

PART II

5. Data and Method

5.1. Research Framework

This chapter will present the research framework, the tools, datasets and methodology used for the analysis of aspect use in advanced Bulgarian and German EFL writing. It will set out with an outline of the general framework and research tools and will proceed with the learner corpus data design, as well as the individual methodological steps employed in the quantitative and qualitative part of the analysis in the present study.

The underlying model for the analysis in the present study is the CIA model proposed by Granger (1996) outlined in the previous chapter – a quantitative and a qualitative comparison between native language and learner language (L1 vs. L2) for the purposes of uncovering distinctive features of "non-nativeness", and between different interlanguages of English (L2 vs. L2) for the purpose of "assess[ing] the generalizability of interlanguage features across learner populations and language situations" (Granger 2009: 18). The model used for the analysis in the present study thus involves the comparison between the native languages and the interlanguages of advanced Bulgarian and German EFL learners, and the comparison of two different interlanguages: the interlanguage of advanced Bulgarian EFL learners with that of advanced German EFL learners. Furthermore, the research framework employed is an expanded form of the CIA framework, since in addition to the two different English interlanguages to be compared and analysed, it also includes several different *native* varieties of English as part of the L1 – L2 comparison: British and American written language produced by novice and expert writers. The reason behind this expanded form of CIA is twofold: on the one hand, it lies in the great variation in the use of the progressive and the perfect in British and American English as two regional varieties of English outlined in chapter 2, and on the other, it lies in the differences between native writing produced by native-speaking writers with little or no experience in expository writing and published writing produced by expert writers. Thus, aspect use in Bulgarian and German learner writing can be compared on the one hand with both British and American use in turn: the aim of this comparison is to establish similarities between a particular interlanguage (e.g. the interlanguage of Bulgarian learners) and a particular target language variety (e.g. British English) as an implicit or an explicit target norm for each EFL learner population. On the other hand, learner writing can be compared with both novice and expert native writing in

order to locate it with respect to a novice – expert native continuum: the aim of this second comparison is to take into account the fact that expository writing and writing competence is a variable and culturally-determined practice (cf. Connor 1996: 16 – 17), which presupposes differences between the rhetorical and organisational patterns of unpublished and published written texts. This continuum is also related to the orality-literacy continuum proposed by Koch and Oesterreicher (1985), which specifies that medially-realised scripturality (e.g. student writing) may still be conceptually very oral in nature (e.g. display higher degrees of conceptual orality, be colloquial in tone), whereas orally-realised speech (e.g. political speeches or university lectures) tends to be conceptually written in nature, even though it is medially spoken (cf. Koch and Oesterreicher 1985 in Günther and Ludwig 1996). Since native speakers of English with little or no experience in writing may consciously or unconsciously "write the way [they] speak!" (cf. Günther and Ludwig 1996: 14) to a much greater extent than experienced native writers, their writing may also show more features of conceptual orality (e.g. greater colloquiality) than the writing of expert writers and thus be closer to the orality end of the orality-literacy continuum. Higher degrees of colloquiality would in turn influence native speakers' use of the progressive and the perfect in both quantitative and qualitative terms (as shown in chapter 2); therefore, a tripartite comparison between learner writing with native novice and native expert British and American writing may help not only to uncover features of "foreign-soundigness" in learner writing, but it may also help to locate learner writing with respect to either a British or an American target language norm, as well as with respect to a native novice-expert literacy continuum.[27] To illustrate, the expanded form of the CIA research framework employed in the present study includes a tripartite comparison between 6 different data sets and is diagrammatically represented in figure 5.1.

---

[27] There are alternative empirical models for the placement of different registers of English along different continua like Biber's multidimensional model (Biber 1988; 1989). Biber uses the co-occurrence of a number of linguistic features to distinguish between different registers of English by grouping them on a scale along five different dimensions: 1) involved vs. informational production, 2) narrative vs. non-narrative discourse, 3) situation-dependent vs. elaborated reference, 4) overt expression vs. argumentation and 5) abstract vs. non-abstract style (cf. Biber in Lüdeling 2008: 834 – 835; Biber 1988; 1989). In the light of the present investigation, it could be supposed that novice native writing would be situated closer to the narrative, non-argumentative, overt and non-abstract end of the scale in comparison expert writing, and it would be interesting to compare the extent to which learner writing resembles either novice or expert writing along each of the five aforementioned dimensions. However, Biber's multidimensional approach goes beyond the scope of the present study, since the linguistic features under investigation (the progressive and the perfect) form only a small part of the complex feature matrix he developed and since they are predominantly relevant to only one of the five dimensions (dimension 2 – narrative discourse).
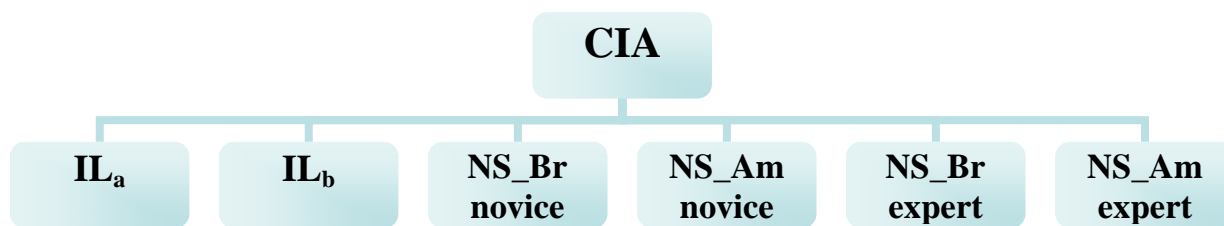
Figure 5.1. Diagrammatic representation of the expanded CIA model used in the present study

Before turning to the corpus material selected for this tripartite comparison, the types of software tools used for the analysis of the learner and native corpora in the present study need to be addressed first.

5.2. Software Tools

Language corpora "by [themselves] can do nothing at all, being nothing other than a store of used language" (Hunston 2002: 3); therefore, they need to be accessed by special corpus software tools which help to rearrange the textual information in them in such a manner that information about frequencies and regularities in language usage can be revealed. The two corpus software tools used for the analysis the present study are *WordSmith* Tools Version 4 (Scott 2010) and *Wmatrix* (Rayson 2008). *WordSmith* is an integrated lexical analysis software tool which analyses how words or patterns of words are used in texts. It has three major functions – *Concord*, *Wordlist* and *Keywords*, of which only the first two will be used for the purposes of the present study. *Concord* is the function which produces concordance lines or concordances showing a particular search word in its immediate linguistic context for the sake of illustrating its meaning in authentic language use (cf. Scott 2004). Concordances are regarded as basic "word-based methods of investigating corpora" (Hunston 2002: 39) that show every instance of a search word (or phrase) in the centre of the computer screen (the so-called *KWIC* format or Key Word in Context) with the respective neighbouring words to the left or right of the search word or phrase. The main use of concordance lines is to "illustrate general and detailed patterns of lexis, word meaning and pattern" (Hunston 2002: 39) that are not necessarily obvious from native-speaker intuition or from a single example of the textual context of the search word. The second *WordSmith* function used for the analysis is the *Wordlist* function (also called "frequency list", cf. Hunston 2002: 67). It generates a list of all words used in a particular text or a corpus and presents them in the order of their frequency of occurrence (and less commonly in an alphabetical order). The purpose of frequency lists is to show the range of vocabulary used in a text or a corpus, as well as to compare the frequency

of individual words across different texts or corpora, which is especially useful when comparing specialized or smaller corpora with general reference corpora and which is not possible on the basis of native-speaker intuitive knowledge (cf. Hunston 2002: 67; Scott 2004). Frequency lists can also be based on more complex linguistic categories like parts of speech, thus allowing for a comparison between the frequencies of nouns, verbs or pronouns across different corpora representing different registers of English; hence, frequency lists of grammatical categories have been employed in various corpus-based reference works and grammars, some of which have been reviewed in the previous chapters (e.g. Biber et al. 1999)[28].

In comparison to *WordSmith*, *Wmatrix* (cf. Rayson 2008) is a web-based corpus tool for corpus analysis and comparison, which can be used for more advanced corpus-linguistic applications (beyond the word form) alongside the standard functions such as wordlists and concordances. *Wmatrix* allows users to upload and run their corpus files through a tagging wizard programme which tags the corpora automatically for parts of speech (POS-tagging), as well as for semantic domains (USAS tags based on the *University Centre for Computer Corpus Research Semantic Analysis System*). In its advanced interface, users can thus compare the frequency profiles generated for different corpora on the basis of words, POS tags and USAS tags, as well as generate concordances for all three types of categories[29]. The *Wmatrix* POS tag frequency lists and the POS tag concordance function have been extensively used in the present analysis for the extraction of frequency lists and concordances for all verb tags and their comparison across the learner and native corpora used in this study. The POS tags are based on CLAWS 7 (the *Constituent Likelihood Automatic Word-tagging System*), which has been continuously developed since the early 1980s and an earlier version of which (CLAWS 4) has been used for the tagging the British National Corpus (BNC) with an average success rate of c. 95% (cf. Garside et al. 1987). On the basis of the 31 different verb tags presented in table 5.2., the verb tag frequencies of the individual corpora can be exported in a Microsoft Word or Excel format, calculated and compared with each other in order to show "whether the analysed word category is equally used, overused or underused by learners and how evenly it is distributed across the [corpora]" (Meunier 1998: 34).

---

[28] For a comprehensive description of WordSmith tools, please visit www.lexically.net

[29] *Wmatrix* also features further applications like n-grams, c-grams and Lemmatiser, which will not be addressed in the present study. For further information, please visit http://ucrel.lancs.ac.uk/wmatrix/

Tag	Category/example
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not... It will be ..)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do... To do...)
VDN	done
VDZ	does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)

Table 5.1. List of CLAWS 7 verb tags (adapted from Rayson 2008)

Automatic comparisons between the POS tags in two different corpora are also possible within the *Wmatrix* interface and can even be graphically represented by the so-called key POS clouds, where statistically significant differences between the frequencies of a particular part of speech are illustrated through font sizes – the bigger fonts stand for more significant differences in statistical terms between the corpora, symbolising significantly over- or underused items[30]. A sample comparison in the form of a 'key POS cloud' between all POS

---

[30] The statistical measure testing keyness and significance employed by *Wmatrix* will be discussed in more detail in section 5.6.2.

tags in the Bulgarian learner corpus and those in the written part of the BNC is presented in figure 5.2.
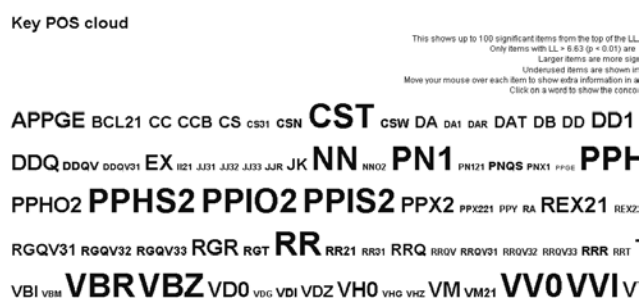
Figure 5.2. Wmatrix key POS cloud of the Bulgarian learner corpus in comparison with the BNC

On the basis of the POS tag frequency lists, all verb tags in a corpus can be extracted and counted automatically, and on the basis of the total verb tag frequencies, various calculations concerning the number of finite (e.g. tensed and modal) or non-finite verb phrases in a corpus can be carried out. The exact procedure of determining the number of progressive and perfect verb phrases will be discussed in greater detail in the methodological section of this chapter (see section 5.6.). In addition, *Wmatrix* can produce concordance lines for each POS tag selected by the researcher, as well as for each semantic USAS tag. However, before turning to the individual methodological steps in the quantitative and qualitative analysis of the present study, the learner and native corpus data design will be outlined in the following two sections.

5.3. Learner Corpus Data Design: *ICLE*

The learner data selected for the present study is based on corpus material from the *International Corpus of Learner English* (*ICLE*), a learner corpus monitored and compiled at the department of corpus linguistics of the University of Louvain-la-Neuve in Belgium (Granger et al. 2002; 2009). *ICLE* is one of the major and most widely-used learner corpora comprising advanced EFL learners' written English produced in an academic setting[31], and one of the very few learner corpora so far which serve as "an empirical resource for large-scale comparative studies of the interlanguage of advanced EFL learners with significantly different native language backgrounds" (Pravec 2002: 83). *ICLE* represents the English of learners who use it as a foreign language (EFL) in the environment of their native language (cf. Gass and Selinker 2001:5), rather than as a second language (ESL) in an English-speaking

---

[31] *ICLE* is also one of the few learner corpora which are publicly available as a CD ROM (cf. Nesselhauf 2006: 141 – 142)

environment or as an official indigenised and nativised variety of English such as Indian English or Nigerian English (cf. Granger 2002: 8) – thus positioning itself within Kachru's Expanding Circle, the Outer Circle being represented by another family of corpora – the *International Corpus of English* (*ICE*) family of corpora[32] (e.g. Kachru et al. 2008). Whereas the first *ICLE* version contains components from 11 different mother-tongue backgrounds (Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Polish, Russian, Spanish, Swedish), the second *ICLE* version (*ICLE* v2) has been upgraded to 16 different mother-tongue backgrounds, including languages from language families other than the Indo-European such as Chinese, Japanese, Turkish and Tswana, and amounting to 3.7 million words in 6,085 essays altogether (cf. Granger et al. 2009: 25). The *ICLE* design criteria rely on a number of shared features and a set of variable features which are illustrated in table 5.3. (for a detailed review of the *ICLE* design criteria, see Granger 1998; 2002; 2003; 2004; 2007; 2009):

Shared features	Variable features
Age	Sex
Learning Context	Mother Tongue
Level	Region
Medium	Other foreign languages
Genre	Practical experience
Technicality	Topic
	Task setting

Table 5.3. *ICLE* design criteria (adapted from Granger 2007: 172)

Among the most important shared variables are the age and the level of the EFL learners – they are defined as "young adults (c. 20 years old)" (Granger 2007: 172) who are advanced learners of English by virtue of the fact that they are at least in their second year of English studies at university, English being their major subject (cf. Granger 2007: 172). Among the most important corpus design differences are the mother-tongue background, the topic and the task setting in which the essays were written – these differences will be further discussed in the present chapter. There are also differences between the first and the second version of *ICLE*: the second version of *ICLE* – *ICLE v2* is not only bigger than the first version, comprising some 6,085 essays and a total number of 3,753,030 words distributed in sixteen national components (cf. Granger et al. 2009: 25), but it is also POS-tagged on the

---

[32] Recent studies on the second version of *ICLE* like Gilquin and Granger (2011) have shown that this dichotomy is not clear-cut, but rather part of an EFL-ESL continuum in the case of *ICLE* v2, where individual *ICLE* components like the Spanish or the French components represent EFL, while others like the Dutch and Tswana components display both ESL and EFL features.

basis of CLAWS 7 and contains a new search interface involving linguistic features, together with user-friendly subcorpus search and compilation options (cf. Granger et al. 2009).

ICLE's potential for interlanguage research has proved to be enormous, not only due to its considerable "balance and systematicity" (Nesselhauf 2006: 148) with regard to its design criteria, but also due to its searchable interface (in particular the second *ICLE v2* version) according to multiple factors that can potentially influence learner language. Among its further advantages for interlanguage research is the comparability of the individual components based on a particular mother-tongue background: they are mostly of the same size (approximately 200,000 words each) and represent student writing which is "fairly neutral, i.e. non-technical and of an average degree of formality" (Nesselhauf 2006: 148). The research goal behind the collection of comparable and tightly-controlled samples of learner writing is to "collect dependable evidence on learners' errors and to compare them cross-linguistically in order to determine whether they are universal or language specific" (Pravec 2002: 83). Nesselhauf (2006: 148) notes that *ICLE* enables

> the researcher to find out whether certain features in the L2 production of a specific L1-group of learners is actually a result of L1 transfer or whether it is a feature more generally present in learner output of a certain target language (and thus potentially a universal feature of L2 production).

ICLE's greatest advantage lies in its potential for cross-linguistic comparisons following the CIA framework (e.g. Granger 1996), where instances of "foreign-soundingness" in the learner essays can be revealed by the "overuse or underuse of words or structures with respect to the target language norm" (Pravec 2002: 83). To ensure comparability between learner and native writing within the CIA research framework (cf. Granger 1996, Gilquin 2008), a target-language usage native control corpus of student writing has been compiled at the University of Louvain-la-Neuve by the name of *LOCNESS* (*Louvain Corpus of Native English Essays*) (cf. Granger et al 2009). *LOCNESS* contains essays written by both British and American high school and university students (although its British part is considerably smaller than the American part and features younger students in their A-levels) who are "novice writers and [whose essays] contain many more errors and infelicities than professional writing" (Granger et al. 2009: 42). Nevertheless, *LOCNESS* provides a good basis for comparison between learner and native writing, since on the one hand, both the *LOCNESS* students and the *ICLE* learners have little or no experience in expository writing in academic settings, and on the other, it allows for a comparison between learner use of a

particular linguistic feature with both British and American use by students at the same level of writing experience[33].

5.4. The Bulgarian and the German Components of *ICLE*

While Granger et al. lament that "up to now, however, learner corpus research has tended to disregard these variables [except for] the influence of the learner's mother tongue" and "the methods used to establish transfer have undeniably often lacked in rigour" (Granger et al. 2009: 45), the present investigation will focus contrastively not only on the mother-tongue background of two different *ICLE* learner populations but will also take into account further factors besides the mere native-language transfer in order to provide for a solid interpretation of the results. For the analysis in the present study, two subcorpora of *ICLE* of approximately equal size have been manually extracted from the Bulgarian and the German components of *ICLE* according to the following two criteria:

(1) the essays had to be written by native speakers of Bulgarian and German
(2) the essays had to be written in response to argumentative essay prompts only

The extracted corpora (which will be called *BUCLE* and *GICLE* for convenience) in terms of their general design, essay codes, institutions, number of essays and number of words are presented in tables 5.4 and 5.5.

Essay codes *BUCLE*	Institution(s)	Number of essays	Number of words	Mean essay length	% whole component
BGSU 1001-BGSU 1302	University of Sofia	300	199,249	664.16	99.5

Table 5.4. *BUCLE* general design

---

[33] *LOCNESS*' design will be outlined in more detail in section 5.5.

Essay codes *GICLE*	Institution(s)[34]	Number of essays	Number of words	Mean essay length	% whole component
DNNI 5008	University of Nijmegen	429	226,503	527.97	96
FRUC 1059 FRUL 1002-	Université Catholique de Louvain				
FRUL 2018	Université de Liège				
GEAU 1001- GEAU 4014	Universität Augsburg				
GEBA 1001- GEBA 1064	Universität Basel				
GEDR 1010- GEDR 1026	Universität Dresden				
GESA 2001- GESA 5045	Paris-London-Universität-Salzburg Wirtschafts- und Fremdsprachenakademie Salzburg				
SWUG 2066	Göteborgs Universitet				
SWUL 8005	Lunds Universitet				

Table 5.5. *GICLE* general design

The extracted corpora *BUCLE* and *GICLE* make up for over 95% of the whole components – the Bulgarian component contains two additional essays written by students with a mother tongue different from Bulgarian, whereas the German component contains 16 additional literary essays amounting to approx. 9,000 words. *BUCLE* and *GICLE* are comparable insofar as they feature argumentative essays written by advanced learners of English in their second or third year of English studies at university level; nevertheless, there are a number of learner- and task-related differences between *BUCLE* and *GICLE*. In terms of their general design, *GICLE* is the bigger and the more diverse learner corpus as it comprises more learner essays coming from a variety of universities mainly across Germany, Austria and Switzerland, whereas *BUCLE* consists of essays written by Bulgarian students of English from only one institution – the University of Sofia in Bulgaria. Hence, *GICLE* is not a German learner corpus in the 'national' sense of the word, but a 'supranational' corpus – its learners will thus be referred to as 'German-speaking' learners in the forthcoming chapters. Moreover, the learner corpora differ significantly with respect to three further factors: the number and character of essay topics and the timing of the essays as two task-related factors

[34] Even though a few of the essays from the German component of ICLE were written in institutions outside of German-speaking countries (i.e. in Belgium and the Netherlands), these are individual essays which were still produced by students with L1 German who were studying in Belgium and the Netherlands at the time ICLE data was collected

and the length of exposure to English in an English-speaking country as an important learner-related factor. Thus, *GICLE* has 83 different essay topics as argumentative prompts, whereas *BUCLE* features only 4 different topics, all of which are covered in *GICLE* too:

(1) In his novel "Animal Farm" George Orwell wrote "All men are equal but some are more equal than others". How true is this today?
(2) Some people say that in our modern world, dominated by science and technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?
(3) A man's/woman's financial reward should be commensurate with his/her contribution to society. Do you agree or disagree?
(4) Most University degrees are theoretical and do not prepare us for the real life. Do you agree or disagree? (cf. Granger et al. 2009)

In addition to these four essay prompts, there are several essay prompts in *GICLE* which are somewhat vague and cannot be classified as strictly argumentative in nature, thus running the risk of encouraging learners to narrate personal stories that happened in the past, rather than argue for or against a particular cause:

(1) Someone I admire!
(2) Telephones
(3) My teenage idol
(4) Fastfood. Yum? (cf. Granger et al. 2009)

In terms of the timing of the essays, all essays in *BUCLE* are untimed, meaning that they were not written in an exam situation under time pressure, whereas only half of the essays (50.1%) of *GICLE* are untimed and almost as much (42% of the essays) were written in an exam situation.[35] The last significant difference refers to the learner-related factor of learners' exposure to English in an English-speaking environment – the percentage of learner exposure in months is illustrated in figure 5.3.

---

[35] There is no exact information on the remaining 8% of the GICLE essays – which fall under the category "unknown"
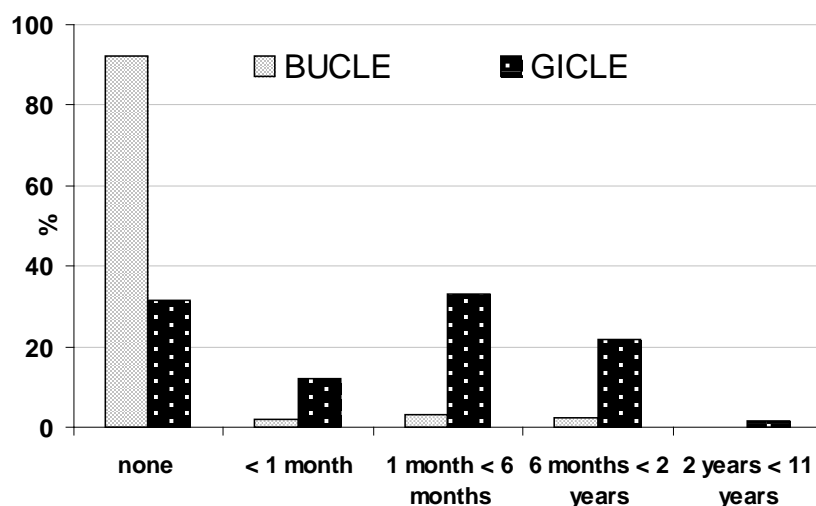
Figure 5.3. Exposure to English in an English-speaking country measured in months

Figure 5.3. demonstrates that the amount of target-language exposure in an English-speaking country is immensely different for *BUCLE* and *GICLE*: whereas over 50% of the German-speaking learners altogether have had at least one month of target-language exposure and the majority even over a month, less than 10% of the Bulgarian learners in *BUCLE* have been abroad at all (cf. Granger et al. 2002; 2009). Furthermore, whereas the majority of the *GICLE* learners have had practical EFL courses taught by native speakers of English (Lorenz 2002: 102 in Granger et al. 2002), the *BUCLE* learners have had very few native speakers as teachers altogether (1 native British or American teacher to 100 students) at school or at university (Blagoeva 2002: 85 in Granger et al. 2002). This suggests that Bulgarian EFL classrooms may be seen to a considerable extent as "impoverished learning environments" (Kasper 1997 in Gilquin and Paquot 2007: 6), providing limited input in the target language in addition to the limited exposure to English in an English-speaking environment[36]. Still, neither the countries of target-language exposure, nor the origin of the native-speaker teachers of the learners in *ICLE* are specified in the learner profiles or in the *ICLE v1* and *v2* handbooks (cf. Granger et al. 2002; 2009), which makes it difficult to categorise the type of target-language exposure for Bulgarian and German EFL learners with certainty.

Another learner-related discrepancy concerning the design of the two learner corpora refers to the distribution between the years of study at university of the Bulgarian and German students – most of the Bulgarian learners are in their second year of studies at the English department of the University of Sofia, whereas the German-speaking learners are more evenly

---

[36] Even though many British and American films, series and sitcoms on Bulgarian TV are not dubbed, the quality, type and amount of general exposure to English for Bulgarian EFL learners can be assumed to be very limited, especially in the late 1990s when the Bulgarian component of *ICLE* was compiled.

distributed across their second, third and fourth year of English studies (20%, 29.4% and 20.5% respectively) (cf. Granger et al. 2002; 2009). This difference is crucial for the proficiency level of the learners – although both Bulgarian and German *ICLE* learners are generically defined as "advanced EFL learners", differences between learners' proficiency levels are bound to occur, especially in terms of the gap between Bulgarian EFL learners in their second year with no exposure to English and German EFL learners of English in their fourth year with over six-month exposure in an English-speaking country. Indeed, Granger and Thewissen (2005) confirm this hypothesis to a certain extent: in an unpublished study on the error-tagged pilot versions of the *ICLE v2* corpora they compare the proficiency levels of the individual *ICLE* components by having twenty randomly selected essays in each component rated by two independent professional raters along the guidelines of the *Common European Framework of Reference* (CEFR) (cf. CEFR, Council of Europe, 1996). The findings for the German and the Bulgarian components for the randomly selected 20 essays (summarised and presented in figure 5.8) show that on average, considerably more German-speaking learners were rated as advanced EFL learners (level C2) than the Bulgarian and all other *ICLE* learners on average, whereas the majority of the Bulgarian learners were rated as upper-intermediate learners (level C1), rather than advanced learners of English, lying thus within the German component – *ICLE* average proficiency range.
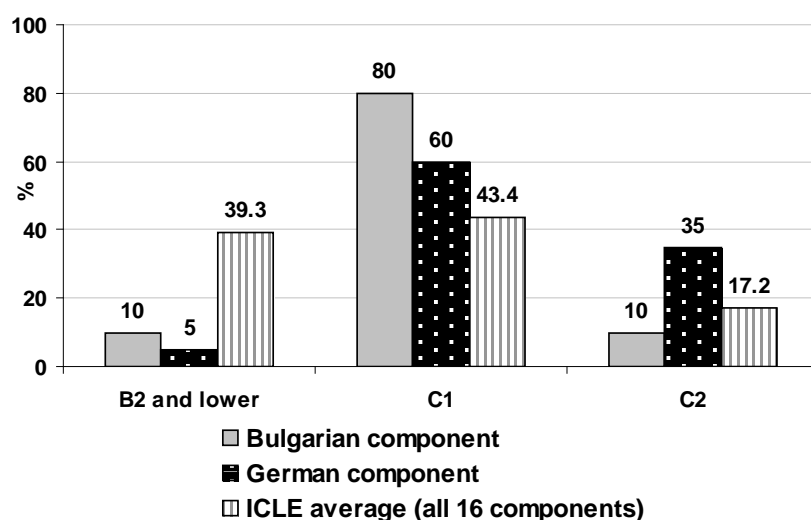
Figure 5.4. A comparison of the independent CEFR ratings for the Bulgarian and German components of *ICLE* with the *ICLE* average (cf. Granger et al. 2009)

Therefore, it can be hypothesised that the Bulgarian and German learners in *ICLE* represent a similar EFL – ESL cline (Bulgarian learners being less advanced and clearly EFL users; German learners more advanced and displaying some ESL uses) to the one investigated by Gilquin and Granger (2011) on the basis of the French, Spanish, Dutch and Tswana

components of *ICLE v2*. Admittedly, the figures for the randomly selected 20 essays represented above might be subject to change if the whole corpora had been rated by professional *CEFR* raters; nevertheless, Gilquin and Granger's (2011) pilot investigation serves as a good point of departure for interlanguage comparisons based on *ICLE*. The learner- and task-related discrepancies between the German and the Bulgarian *ICLE* components and the resulting differences in aspect use will be further discussed in chapters 6 and 7 in the present study.

To summarise, despite the learner- and task-based variation between the individual *ICLE* components and *BUCLE* and *GICLE* in particular, *ICLE*'s potential for interlanguage research still remains enormous with regard to its design criteria, searchability and uniqueness as one of the few major academic learner corpora which have been more or less tightly controlled in the process of their compilation. The next section will present the design of the native-speaker corpora used for the analysis of the present study.

5.5. Native Corpus Data Design: *LOCNESS*, *FLOB* and *FROWN*

The control corpora used for this study are based on material from four different corpus sets – *ICLE*'s 'sister' corpus of native-speaker essays *LOCNESS*, split into a British and an American part (called for convenience *LOCNESS_br* and *LOCNESS_us*) and the F sections of the *FLOB* and *FROWN* corpora of British and American English (called for convenience *FLOB_F* and *FROWN_F*) (cf. Granger et al. 2002; 2009; Hundt et al. 1998; 1999). These four different control corpora have been selected in view of the aforementioned considerations with regard to the differences between novice and expert native writing, as well as with regard to the quantitative and qualitative differences between the use of the progressive and the perfect in the two major varieties of English – British English and American English outlined in chapter 2. Notably, comparing three different corpus sets – learner writing with native novice writing and native expert writing ensures a comprehensive description of learner variation along the native novice-expert continuum, as well as with respect to learners' possible influence by either British or American English target norms with regard to aspect use. The six corpora involved in the tripartite comparison presented in section 5.1. of the present chapter are illustrated in figure 5.5.
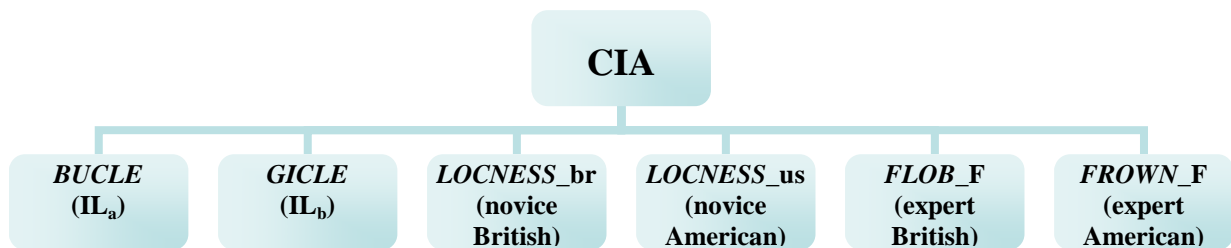
Figure 5.5. Learner and native corpus data sets in the present study

The first native-speaker corpus, *ICLE*'s 'sister' corpus *LOCNESS* (*Louvain Corpus of Native English Essays*) is a corpus consisting of native-speaker essays written by British and American high school and university students, which was monitored and compiled at the University of Louvain to ensure comparability with *ICLE* within the Contrastive Interlanguage Analysis framework (cf. Granger 1996; Granger et al. 2002; 2009). For the purposes of the present study and in view of the differences in aspect use in British and American English, *LOCNESS* has been split into two parts – a British part – *LOCNESS_br* and an American part – *LOCNESS_us*. *LOCNESS_br* is the smaller corpus of the two novice-writer corpora (comprising approximately 80,000 words) and the only one which contains a substantial portion of essays written by very young adults (A-level high-school students in Britain), whereas its American counterpart *LOCNESS_us* is almost twice as big (c. 150,000 words) and considerably more diverse, featuring essays collected in 5 different universities across the United States. The general design of *LOCNESS_br* and *LOCNESS_us* is presented in tables 5.6 and 5.7.

Essay codes *LOCNESS*_br	Institution(s)	Number of texts	Number of words	Mean text length
ICLE-ALEV-0001/10.6- *ICLE*-ALEV-0001/139 *ICLE*-BR-SUR-0001.3- *ICLE*-BR-SUR-0033.3	Unknown: school and university students (A levels)	147	79,228	539

Table 5.6. *LOCNESS_*br general design

Essay codes LOCNESS_us	Institution(s)	Number of texts	Number of words	Mean text length
ICLE-US-MRQ-0001.1-*ICLE*-US-MRQ-46.1	Marquette University	176	149,573	849.8
ICLE-US-IND-0001.1-*ICLE*-US-IND-28.1	Indiana University at Indianapolis			
ICLE-US-PRB-0034.2-*ICLE*-US-PRB-39.2	Presbyterian College, South Carolina			
ICLE-US-SCU-0001.1-*ICLE*-US-SCU-17.4	University of South Carolina			
ICLE-US-MICH- *ICLE*-US-MICH-0001.1-45.1	University of Michigan			

Table 5.7. *LOCNESS*_us general design

The selection of the native-speaker essays was subject to the same restrictions as the selection of the learner essays: in order to guarantee the best possible comparability with *BUCLE* and *GICLE*, only argumentative essays from both *LOCNESS* corpora have been selected for the present study (which accounted for the majority of the essays). Still, similar to the essay topics in *GICLE*, many of the essay topics in both *LOCNESS_br* and *LOCNESS_us* could not be classified as clearly argumentative and well-formulated: thus, essay prompts like "BSE and British beef" or "Fox hunting" in *LOCNESS_br* could be classified as potentially encouraging learners to narrate personal stories or describe situations that happened in the past rather than argue for or against the topic in question. The same limitation applies to *LOCNESS_us* which features a wide variety of essay topics (over 40 different topics), many of which are superficially argumentative in nature, but still rather vaguely put and subject to misinterpretation by American students (e.g. "Violence on television", "Legalization of marijuana" or "Capital punishment") (cf. Granger et al. 2002; 2009).

The remaining two control corpora – *FLOB* and *FROWN* – have been selected as representatives of expert British and American writing due to the fact that they represent written material that has been published either in Britain (*FLOB*) or in the United States (*FROWN*) (cf. Hundt et al. 1998; 1999). *FLOB* and *FROWN* are the 1990's Freiburg updates of the well-known and widely-used *BROWN* family of corpora of native-speaker writing, which were compiled and published as standard reference corpora of British and American written English in the 1960s and which consist of 5000 different samples of approximately

2,000 words each (cf. Francis and Kucera 1979). The sampling criteria of *FLOB* and *FROWN* follow closely the compilation of their predecessors *BROWN* and *LOB* on the basis of a semi-random principle, which does not capture representative samples in a strict statistical sense (cf. Hundt et al. 1998; 1999), but which matches the *LOB/BROWN* corpus material as closely as possible by sampling the same newspapers, journals, books and periodicals that were used for the original corpora *BROWN* and *LOB*. The F sections of *FLOB* and *FROWN* (called for convenience *FLOB_F* and *FROWN_F*) selected for the present study represent British and American non-fictional, non-technical popular writing termed 'popular lore' in the corpus manuals (cf. Francis and Kucera 1979; Hundt et al. 1998; 1999). The material for the F sections was extracted from various magazines and journals like e.g. *Family Circle, Elle, National Review* etc., as well as non-fiction books and how-to guides on a variety of popular topics (cf. Hundt et al. 1998; 1999).

These sections of *FLOB* and *FROWN* have been selected as comparable control corpora in the tripartite comparison between learner and native writing due to their level of non-technicality together with their higher degree of argumentativeness in contrast to the other *FLOB* and *FROWN* non-fiction sections which comprise news reportage, religious, technical or scientific articles. An additional advantage of the F sections of *FLOB* and *FROWN* lies in the fact that they feature mostly commentaries on popular topics related to the world affairs in the 1990s and are thus similar to the learner and native novice corpora both in terms of their topic choice being "current affairs" and in terms of linguistic criteria like their tense choice being predominantly the simple present[37]. The general design of *FLOB_F* and *FROWN_F* is illustrated in table 5.8. and 5.9. (cf. Hundt et al. 1998; Hundt 1999 for a detailed description of the design of the F sections).

Essay codes *FLOB_F*	Source	Number of texts	Number of words	Mean text length
F01-F44	British publications: books and articles	44	88,574	2013

Table 5.8. *FLOB*_F general design

---

[37] The remaining non-fiction sections of *FLOB* and *FROWN* have not been selected as control corpora for the present investigation either because of their level of technicality (e.g. religion, trades and hobbies in sections D and E) and target readership (the three press sections A, B and C), or because of the design of the samples including narrative subsections (e.g. sections G – essays, *belles lettres* and biographies) and thus having a direct influence on the simple present – simple past ratio in the respective corpus sections.

Essay codes *FROWN*_F	Source	Number of texts	Number of words	Mean text length
F01-F48	American publications: books and articles	48	96,587	2012

Table 5.9. *FROWN*_F general design

The next two sections will address the individual methodological steps in the quantitative and qualitative analysis carried out in the empirical part of the present investigation.

5.6. Methodology I: Quantitative Analysis

The methodology used in the present study is a mixture of quantitative and qualitative research methods (the so-called mixed research methods, cf. Dörnyei (2007: 24) which aims to "bring out the best of both paradigms, thereby combining quantitative and qualitative research strengths" (Dörnyei 2007: 45). The quantitative research strengths lie in the use of numbers and predefined categories to standardise and objectify research procedures and arrive at generalisable and ideally universal results (ibid.: 33-34), whereas qualitative research methods are used to make sense of complex phenomena with the help of longitudinal examinations and interpretations of the results (ibid.: 39-40). Both research methods have weaknesses, ranging from the overall insensitivity of quantitative research methods to the context and dynamics of the observed phenomena to the lack of methodological rigour and sample generalisability of qualitative investigations (ibid.: 33-41). In contrast, mixed research methods increases the strengths of the two approaches while minimising their weaknesses by involving a multi-level analysis of complex issues which includes a quantitative phase setting numeric trends, followed by a qualitative phase which focuses on specific details to explain the observed quantitative trends and tendencies (ibid.: 45). The present study will include both a quantitative corpus-based phase in the analysis of learner and native use of aspect in writing, as well as a subsequent qualitative interpretation of the results obtained by the quantitative analysis. The individual steps in both phases will be outlined next.

5.6.1. Frequency Measurement and Retrieval of Aspect Forms

Since the frequency of words and other linguistic units lies at the heart of corpus-based methodology (e.g. Hunston 2002), thus making it different from all other approaches to language (cf. Baroni 2009: 803), this section will first set out to define the retrieval and calculation of the frequencies' of the two aspect constructions as part of the quantitative analysis procedure before coming to the qualitative analysis in 4.5. To this end, all six corpora – the two learner corpora *BUCLE* and *GICLE* and the four native corpora (*LOCNESS_br*, *LOCNESS_us*, *FLOB_F* and *FROWN_F*) were uploaded and run through the web-based tag wizard of *Wmatrix* for the purpose of tagging them for parts of speech (POS) (cf. Rayson 2008). Subsequently, frequency lists for all POS-tags have been produced by the *Wmatrix* tag wizard and all verbal tags have been automatically filtered from the general POS-tag frequency lists and extracted for each corpus individually, including both the absolute verb tag frequencies and the relative verb tag proportions in percent of all POS-tags. A sample verb tag frequency list automatically extracted for the learner corpus *BUCLE* is presented in table 5.10[38]. This table illustrates nicely that e.g. the POStags with the highest frequencies in *BUCLE* are – as expected – the verb *be* in its present tense form (singular – VBZ and plural – VBR), alongside with lexical verbs in the simple present (VV0 and VVZ), non-finite lexical verbs (VVI) and modal verbs (VM).

V tag	#	%	V tag	#	%	V tag	#	%	V tag	#	%
VB0	1	0	VBZ	4594	2.42	VHD	173	0.09	VV0	4485	2.36
VBDR	265	0.14	VD0	566	0.3	VHG	104	0.05	VVD	934	0.49
VBDZ	389	0.2	VDD	84	0.04	VHI	379	0.2	VVG	2643	1.39
VBG	197	0.1	VDG	50	0.03	VHN	23	0.01	VVGK	22	0.01
VBI	1791	0.94	VDI	186	0.1	VHZ	691	0.36	VVI	6706	3.53
VBM	173	0.09	VDN	56	0.03	VM	3736	1.97	VVN	3849	2.03
VBN	347	0.18	VDZ	236	0.12	VM21	42	0.02	VVNK	6	0
VBR	2531	1.33	VH0	1172	0.62	VMK	20	0.01	VVZ	2286	1.2

Table 5.10. Verb POS tag frequency list for *BUCLE*

---

[38] Granger (2002: 18) comments on the problems of applying automatic POS-taggers to learner data; however, she notes that POS-taggers can tag advanced learner writing featuring few spelling and morphological errors with similar success rate to native-speaker data and laments that few learner corpus studies utilise POS-tagged learner corpora (cf. Aarts and Granger 1998; Granger and Rayson 1998)

In order to retrieve the frequencies of the progressive and the perfect verb phrases on the basis of the verb tag frequencies calculated by *Wmatrix*, a number of relevant verb tags were selected for a manual refinement and further analysis. In order to retrieve the progressive verb phrases in all six corpora, all verb tags representing the *–ing* suffix were selected for concordancing and the corresponding concordance lines were exported from *Wmatrix* into Excel spreadsheets. The extracted verb tags were VBG, VDG, VHG and VVG, standing respectively for the verbs 'being', 'doing', 'having' and all other lexical verbs which were potential candidates for the non-finite participle slot of the progressive. Next, all *–ing* forms functioning as non-progressives (e.g. deverbal adjectives, adverbial participles, gerunds, nominalisations and non-finite relative clauses, cf. König and Gast 2009: 72) were manually filtered out and deleted from the retrieved concordance lines. The remaining verb phrases were carefully classified as progressives by looking at the contexts they occurred in; progressives of the future *going-to* type were also discarded from the concordance lines on account of their non-aspectual function. Furthermore, in order to improve precision, especially with regard to the learner data, WordSmith concordances (cf. Scott 2004) were separately run (i.e. as a "double check") for all words ending in *–ing*, and the frequency results obtained after the manual refinement of the *Wmatrix* frequency lists for the four verb tags VBG, VDG, VHG and VVG were carefully matched with the results obtained by the word search with *WordSmith*. Those progressive verb phrases obtained by *WordSmith* that were missing in the refined *Wmatrix* frequency lists were added to the count.

The perfect verb phrases were retrieved on the basis of *Wmatrix* concordances of the finite present and past auxiliary verb tags VH0 'have', VHD 'had' and VHZ 'has' followed by anything and were manually filtered by deleting all non-perfect forms of 'have' as a full verb followed by an object or a complement. In addition, all perfect verb phrases of the type 'have got' followed by an object were also manually deleted on account of their non-aspectual non-perfect function. Finally, all modal progressives and perfects were discarded from the count, since they fall under the category 'modality' and are therefore not part of the present analysis which focuses on grammatical aspect exclusively. The filtered concordance lines of the progressive and perfect VPs were exported into Excel spreadsheets and the type-token ratios of lexical verbs in the progressive and perfect were calculated. Next, the raw frequencies of the progressive and the perfect verb phrases in the six corpora retrieved via the *Wmatrix* online tag concordancing and refined by the *WordSmith* concordances were normalised to relative frequencies per thousand words by taking into consideration the size of the respective corpora, thus making the frequencies of the progressive and the perfect directly

comparable with the frequencies obtained in the corpus-based studies on first and second-language aspect use reviewed in chapters 2 and 3 (e.g. Axelsson and Hahn 2001; Biber et al. 1999; Elsness 2009; Erikson 2008; Granger 1999; Mindt 2000; Schlüter 2002; Smitterberg 2005 etc.). Thus, the first type of frequency measurement of aspect forms used in the present study is the normalised frequency, calculated with the help of the following formula:

$$\textbf{Normalised frequency aspect form} = \frac{\textbf{N aspect form occurrences}}{\textbf{N tokens in a corpus}} * \textbf{1000}$$

In addition to the normalised frequencies per 1,000 words of the progressive and the perfect, the present study utilises another type of frequency measurement which compares the frequency of the progressive and the perfect measured in relation to the number of finite verb phrases in a corpus. This type of measurement reflects the fact that corpora with the same number of words (tokens) may have different shares of verb phrases in relation to other parts of speech – i.e. some corpora (e.g. corpora featuring spoken language) may be 'verbier', whereas others tend to be 'nounier' (e.g. corpora containing written language). Therefore, the widely-used normalised frequencies per 1,000 words reviewed in the previous chapters are not always appropriate for measuring the frequencies of grammatical phenomena, since they do not reflect the fact that a grammatical feature like the progressive or the perfect is not just a substitute for any word in English, but only for non-progressive or non-perfect verb phrases (cf. Smitterberg 2005: 40 – 44).

Hence, the present study utilises a new type of frequency measurement that has not been employed to this extent[39] by corpus-based studies on grammar so far, which takes into consideration the number of verb phrases in a corpus and counts the frequency of the progressive and the perfect relative to this number. This measurement is similar to the so-called V-coefficient used for the calculation of the progressive outlined by Smitterberg in his study on the distribution of the progressive in 19th century English (2005: 44 – 45): the V-coefficient "relates the number of progressives to the number of verb phrases, thus neutralizing the differences between samples in ratios of the number of verb phrases to the number of words" Smitterberg (2005: 44); it is easy to calculate in a tagged corpus and is better suited as input to statistical tests of comparison measuring variation in the distribution of progressive and non-progressive verb phrases across different corpus sets like the chi-square test (cf. Smitterberg 2005: 44 – 45). Still, the V-coefficient does not specify whether a

---

[39] with the exception of Smitterberg (2005)

verb phrase can potentially be a progressive or a perfect verb phrase; therefore, for the purposes of the present study, the V-coefficient has been adapted to include only finite verb phrases (tensed and modal verb phrases) which can potentially be progressive or perfect[40]. Since all six corpora have been tagged by Wmatrix on the basis of CLAWS 7, the number of finite verb phrases could be determined semi-automatically, following the finite verb phrase models suggested by Halliday and James (1993) and Quirk et al. (1974). Halliday and James (1993) define finite clauses as "verbal groups which embody a choice of deixis […], [where] verbal deixis is either (a) modality or (b) primary tense" (Halliday and James 1993: 39). Counting the instances of the temporal and modal finite operators in a corpus is thus helpful for identifying the number of finite verb clauses, which entails the number of finite verb phrases (cf. Halliday and James 1993: 39 – 40). In a similar way, Quirk et al. (1974: 73) identify four basic types of finite verb phrases in their *Grammar of Contemporary English*:

(a) **Type A (Modal/periphrastic)** – consists of a modal or periphrastic auxiliary + the base of the verbal-phrase head. For example: He *must examine*.
(b) **Type B (Perfective)** – consists of the auxiliary *have* + the *–ed* participle of the verb-phrase head. For example: He *has examined*.
(c) **Type C (Progressive)** – consists of the auxiliary *be* + the *–ing* participle of the verb-phrase head. For example: He *is examining*.
(d) **Type D (Passive)** – consists of the auxiliary *be* + the *–ed* participle of the verb-phrase head. For example: He *is examined*.

The number of finite verb phrases in the six corpora was thus carefully calculated in accordance with these two finite verb phrase models – by adding the frequencies of the modal and temporal finite operators signalled by the following selected CLAWS7 verb tags: VBDR, VBDZ (*were*, *was*), VBM (*am*), VBR, VBZ (*are*, *is*), VD0 (*do*, finite base form), VDD (*did*), VH0 (*have*, finite base form), VHD (*had*), VHZ (*has*), VM (modal), VMK (modal catenative), VV0 (lexical verb), VVD (lexical verb, past tense), VVZ (lexical verb, third person singular). In this way, the overall frequencies of the finite verb phrases in the corpora could also be compared with each other along the lines of the CIA framework; furthermore, the frequencies of the individual tensed and modal verb phrases could be deducted from the overall frequencies of the finite verb tags and compared across the six corpora. Simple present and simple past verb phrases were thus deducted from the number of finite present and past operators (e.g. VBM, VBR, VBZ, VD0, VH0, VHZ, VV0, and VVZ for the simple present) by summing up the frequencies of the relevant finite operators and subtracting the frequencies

---

[40] The V-coefficient is still a very general measurement, since there are a number of restrictions on the use of the progressive and the perfect (cf. chapters xyz) and many of the finite verb phrases cannot be either progressive or perfect; nevertheless, it helps to refine precision by calculating proportions of all verb phrases rather than normalising frequencies to a number of tokens.

of the individual finite operators which were part of non-simple aspect constructions (e.g. by subtracting from the general *Wmatrix* VH0 number the number of VH0 tags which function as a present perfect plural marker). Hence, the number of finite verb phrases was subdivided into simple present, simple past, present and past perfect, present and past progressive, as well as modal verb phrases, and the frequency of each of those was further normalised as a proportion of all finite verb phrases in percent, as well as in relation to the number of words in the six corpora (per 1,000 words). The distribution of the finite verb phrases is diagrammatically represented in figure 5.6.
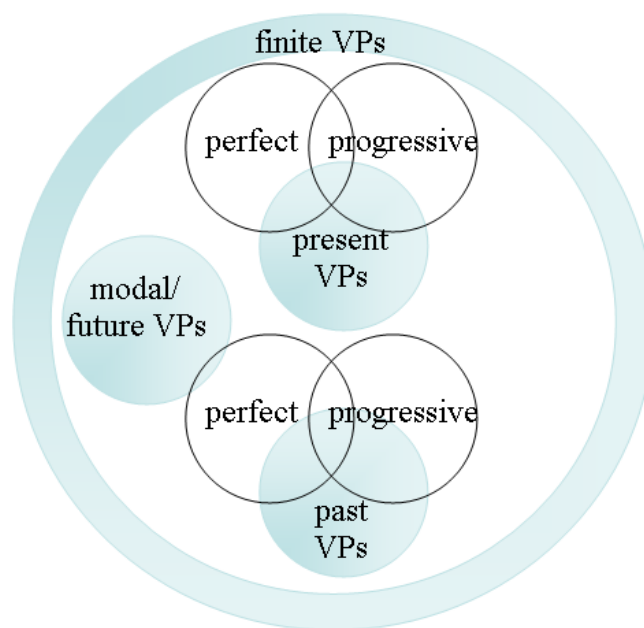
Figure 5.6. Diagrammatic representation of the subsets of finite verb phrases

Hence, the second type of frequency measurement employed in the calculation of aspect forms in the present study is the frequency proportion in %, calculated with the help of the following formula:

$$\text{Frequency proportion aspect form (in \%)} = \frac{\text{N aspect form occurrences} * 100}{\text{N finite verb phrases}}$$

In addition to the retrieval and frequency calculation of progressive and perfect VPs, the frequency of temporal adverbials accompanying them was manually retrieved from the concordance lines of the progressive and the perfect, the goal being a comparison of the degree of temporal modification (especially with the stereotypical 'trigger word' adverbials, see chapter 2) across the two learner and four native corpora. The number of contracted present and past auxiliaries occurring with the progressive and the perfect (e.g. *they're doing; he's done* etc.) was also calculated in view of the tendency of EFL learners of various mother-

tongue backgrounds to adopt features "that are more typical of speech than of academic writing" (Gilquin and Paquot 2007: 10) and that are stylistically inappropriate for the genre 'argumentative essay', being closer to the orality end of the orality-scripturality cline. Finally, the distribution of progressive and perfect VPs across main and subordinate clauses was calculated in relation to the claims of the Discourse Hypothesis (cf. chapter 3) that perfective verb phrases are often found in foregrounded or main clauses and progressive verb phrases in backgrounded or subordinate clauses which render conditions, accompanying circumstances, descriptions and actions[41]. Calculating the frequencies of temporal adverbials and auxiliary contractions accompanying the progressive and the perfect and determining the type of clauses they occur in allows for an in-depth quantitative comparison of learner and native writing with respect to the factors which govern the use and distribution of these two aspect constructions in native and non-native English. The means of comparison between the learner and native corpora and the tests proving statistical significance will be outlined briefly in the next section.

5.6.2. Frequency Comparison of Aspect Forms and Statistical Tests

Since corpus-based techniques are commonly used to examine variation in language usage across different data sets representing different genres or users (cf. Rayson, Berridge and Francis 2004: 926), a frequency comparison of words or other linguistic items in the different corpora is the starting point of every such examination. Rayson and Garside (2000: 1) and Rayson, Berridge and Francis (2004: 926 – 927) define two basic types of comparison between corpora:

(1) A comparison of a sample corpus with a large(r) standard corpus (e.g. Scott 2004)
(2) A comparison between two (roughly-) equal sized corpora (e.g. Granger 1998)

The second type of comparison views corpora "as equals. It aims to discover features in the corpora that distinguish one from another" (Rayson, Berridge and Francis 2004: 927) and is the type of comparison which lies at the heart of the quantitative analysis in the present study. Several issues need to be considered before such a comparison is carried out: the comparability of the corpora and the homogeneity within them, as well as the reliability of

---

[41] Contrary to the claims of the Discourse Hypothesis with regard to the distribution of the progressive predominantly in backgrounded or subordinate clauses, recent corpus-based grammars (cf. chapter 2) have proven that both the progressive and the perfect occur predominantly in main clauses

statistical tests proving significant frequency differences between them (cf. Rayson and Garside 2000:1; Rayson, Berridge and Francis 2004: 927). Whereas the learner corpora *BUCLE* and *GICLE*, as well as the native control corpora *FLOB* and *FROWN* have been designed as more or less homogeneous counterparts to each other (*BUCLE* to *GICLE* and *FLOB* to *FROWN* respectively), mainly in terms of size and sampling criteria[42], the British and American parts of *LOCNESS* (split for the purposes of the present study) are slightly less comparable in size, *LOCNESS*_us being much bigger than *LOCNESS*_br. To ensure reliability of statistical tests in relation to the different-sized corpora in the present study, Dunning's (Dunning 1993 in Rayson and Garside 2000: 2) log-likelihood ratio G^2 (henceforth also LL statistics) has been applied to all comparisons between the learner and native frequencies of use. The log-likelihood ratio is a goodness-of-fit statistical model which compares the observed and expected frequencies of words or other linguistic items across two corpora in order to determine significant deviations in one corpus (a normative corpus) in comparison to a comparative corpus (cf. Rayson, Berridge and Francis 2004: 928 – 929). It can be applied to comparisons between words, POS tags or semantic tags across different corpora and can be used "to discover key items in the corpora which differentiate one corpus from another" (Rayson and Garside 2000: 5)[43]. The LL statistics is performed with a contingency table like table 5.11, where for a comparison on the lexical level the values 'a' and 'b' correspond to the frequencies of a particular word in the two corpora, and the values 'c' and 'd' correspond to the number of words in total.

	Corpus one	Corpus two	TOTAL
Freq of word	a	b	a+b
Freq of other words	c-a	d-b	c+d-a-b
TOTAL	c	d	c+d

Table 5.11. Contingency table for word frequencies (adapted from Rayson and Garside 2000: 3)

The values 'a' and 'b' are the observed values (O), the 'c' and 'd' values correspond to the N values in the following formula used for the calculation of the expected values (E) (cf. Rayson and Garside 2000: 3):

The expected word frequencies E1 and E2 are calculated by considering the number of words in the corpora with the help of the two formulae (ibid.: 3):

E1 = c*(a+b)/(c+d) and E2 = d*(a+b)/(c+d)

The LL ratio itself is then calculated with the formula (ibid.: 3):

$$-2\ln \lambda = 2\sum_i O_i \ln\left(\frac{O_i}{E_i}\right)$$

The last formula equates to calculating the LL as follows ((ibid.: 3):

$G^2 = 2*((a*\ln (a/E1)) + (b*\ln (b/E2)))$

Higher LL values correspond to more significant differences between the relative frequencies in the two corpora which need further investigation and qualitative description in order to establish practical significance of the results and "make hypothesis about the corpora and the language use they represent" (Rayson and Garside 2000: 5). The two learner and four native corpora in the present study will thus be compared and frequency-profiled in pairs in terms of their frequency of use of aspect forms and other linguistic items influencing the use of aspect with the help of the LL statistics[44]. The last methodological steps of the present analysis involve a qualitative examination of the learner and native uses of aspect in English and will be presented in the next section.

---

[44] All statistical calculations use the LL Excel downloadable spreadsheet developed by Rayson, see http://ucrel.lancs.ac.uk/llwizard.html

5.7. Methodology II: Qualitative Analysis

Although all corpus-based research is intrinsically quantitative and relies on the frequency of use of words and other linguistic items (cf. Hunston 2002: 3 – 4), a qualitative evaluation of the results is still necessary for the appropriate description and interpretation of the observed phenomena – this is where "the researcher must intervene and qualitatively examine examples of the significant words highlighted by the [quantitative] technique" (Rayson and Garside 2000: 3). The final part of this chapter will outline those steps in the analysis which allow for a qualitative interpretation of the results obtained by the corpus-based methodology presented in the previous sections. As a first step towards the qualitative evaluation of aspect use in native and learner writing, all progressive and perfect verb phrases extracted from the six corpora were manually assigned one of the four inherent lexical verb classes after Vendler (1957): states, activities, accomplishments and achievements by means of diagnostic tests for lexical aspect. Since lexical aspect is a compositional property of verb phrases as entities and includes their nominal arguments (e.g. objects and complements), all progressive and perfect verb phrases were carefully examined in their larger context with regard to inherent semantic properties like telicity or punctuality (cf. Haznedar 2007: 391). In this way, an across-category analysis which compares the percentage of all progressives that are activities or all telic perfects across the learner and native corpora can be carried out in order to prove the claims of the Aspect Hypothesis for advanced EFL learners' argumentative writing (cf. Bardovi-Harlig 2000; 2002, see chapter 2). The diagnostic tests used for distinguishing between the four different classes of lexical verbs (states, activities, accomplishments and achievements) were based on several influential theoretical accounts on the inherent lexical aspect of verb phrases such as the classifications found in Vendler (1957), Dowty (1979: 66 – 71), Andersen and Shirai (1995: 749) and Brinton (Brinton 1998: 242; Brinton 2000: 143 – 147).

Two major diagnostic tests (tests a) and b), see tables 5.12 and 5.13) for determining the inherent lexical aspect of the progressive and the perfect verb phrases were adopted for the purposes of the present study: Andersen and Shirai's (1995) and Brinton's (1998) diagnostic tests for inherent lexical aspect:

Step 1	**state or non-state** **Does it have a habitual interpretation in the simple present tense?** **If no → state (e.g. *I love you*)** **If yes → non-state (e.g. I eat bread) → go to step 2**
Step 2	**activity** or **non-activity** Does *X is Ving* entail *X has Ved*? Without an iterative/habitual meaning? In other words, if you stop in the middle of *Ving*, have you done the act of *V*? If yes → activity (e.g. *run*) If no → non-activity (e.g. *run a mile*) → go to step 3
Step 3	**accomplishment** or **achievement** If test a) does not work, apply test b) and possibly c) a) If *X Ved in Y time* (e.g. 10 minutes), then X *was Ving* during that time if yes → accomplishment (e.g. *He painted a picture*) if no → achievement (e.g. *He noticed a picture*) b) Is there ambiguity with *almost*? If yes → accomplishment (e.g. *He almost painted a picture* has 2 readings → *He almost started painting a picture* or *He almost finished painting a picture*) If no → achievement (e.g. He almost noticed a picture has only one reading) c) *X will VP in Y time* (e.g. 10 minutes) = *X will VP after Y time* If no → accomplishment (e.g. *He will paint a picture in an hour* is different from *He will paint a picture after an hour*) If yes → achievement (e.g. *He will start singing in two minutes* is the same as *He will start singing after two minutes*)

Table 5.12. Diagnostic test a) for lexical aspect (adapted from Andersen and Shirai 1995: 749)

Activities	consist of successive phases answer the question 'For how long?' go on for a time in a homogeneous way; do not take any definite time cannot be finished if one stops *Ving*, then one did *V* call for periods of time that are not unique or definite
Accomplishments	answer the question 'How long did it take?' have a terminal point or 'climax' which is logically necessary take a certain time and do not go in a homogeneous way if one stops *Ving*, then one did not *V* if one *Vs* in an hour, then at any time during the hour one is *Ving* imply the notion of unique and definite time periods
Achievements	are predicated for a single moment answer the question 'At what time?' some are not actions as soon as one *Vs*, one has *Ved* if one takes an hour to *V*, then at any time during the hour one is not *Ving* involve unique and definite time instances
States	are predicated for a given period of time answer the question 'For how long?' often name abilities, qualities, habits do not occur with *deliberately*, *carefully* involve time instances that are indefinite and non-unique

Table 5.13. Diagnostic tests b) for lexical aspect (adapted from Brinton 1988: 242)

The results obtained with the help of the diagnostic tests illustrated above were carefully matched against more recent classifications presented in a number of second language acquisition studies featuring verb lists such as e.g. Bardovi-Harlig and Reynolds (1995), Collins (2002) and Haznedar (2007)[45]. A sample summarised classification of verbs according to their inherent lexical aspect is presented in table 5.14 (based on Brinton 1988: 241-243; Brinton 2000: 144-147; Collins 2002: 94; and Dowty 1979: 66-71) – the verbs in bold represents those contentious cases which belong to several aspectual verb classes according to the different authors).

---

[45] There are certain disparities in the more recent empirical studies concerning e.g. the inherent lexical aspect of telic verbs and their exact classification as either accomplishments or achievements; however, such a fine-grained differentiation between accomplishments and achievements will not play a role for the analysis in the present study since it does not influence the results with respect to the claims of the Aspect hypothesis concerning the use of the progressive and the perfect – thus both accomplishments and achievements will be grouped as "telic verbs" or "telics".

states		activities		accomplishments		achievements	
astonish	love	**attend**	roll	**attend**	obliterate	arrive	**hear**
be	**look**	continue	rotate	appoint	paint sth	awaken	improve
believe	mean	**cook**	rumble	so	perform sth	**be** born	**kill**
belong	need	cry	**run**	box	place sth	become	**know**
concern	own	dance	scan	buy	**play** (a	begin	land
desire	perceive	drink	scrutinize	bring	game)	break	leave
dislike	possess	**drive**	search	(about)	put	catch	lose
dismay	prove	eat	seek	build sth	recover	cease	melt
doubt	rule	gaze upon	sing	cause sth	(from an	cool	notice
dominate	**regret**	focus on	sit (in/on)	VP	illness)	(down)	reach (the
enjoy	**see**	follow (with	smile	change	read (a	cross (the	summit)
exist	seem	the eyes)	smoke	(the	book)	border)	**see**
feel	show	**go**/attend	stay	story)	rent sth	darken	start
hate	suggest	housekeep	study	cover	**ride** (10 km)	depart	**taste**
have	smell	hum	**swim**	**cook** sth	**run** (5 km)	detect	**think** of
hear	taste	keep	talk	deliver	**run** (away)	die	touch
imply	think that	**listen** (to)	**tell**	sth	**swim** (5 km)	discover	**turn** off
involve	**understand**	**look** (for)	(stories)	destroy	see Carmen	drop	**turn** into
know	want	observe	**think**	draw sth	set sth	end	spot sth
like	worry	panic	(about)	**drive** (to	shape up	explode	realise
		pay	vibrate	X)	take (out)	fall (out)	recognise
		(attention	**walk**	fly (to	**tell** (a story)	feel	remember
		play	**watch**	X)	**turn** sth into	find	resume *Ving*
		push sth	work	**get**	sth	finish	**see**
		pull sth	**write**	exhauste	uncover	forget	sink
		ride (on)	(in/on)	d	**write** sth	freeze	start *Ving*
				get ready	**walk** (to)	**get**	stop *Ving*
				give	**watch** sth	married	**understand**
				go to		happen	warm (up)
				(Paris)			win
				go (out)			
				grow up			
				hide			
				kill			
				knit sth			
				make sth			
				VP			
				marry			
				move			

Table 5.14. Sample list of verbs according to their inherent lexical aspect (contentious cases in bold)

As a final step in the qualitative part of the methodology, learner uses of the progressive and the perfect aspect were evaluated by a native informant – a speaker of American English, freelance journalist and expert writer in terms of their acceptability within the temporal framework and discourse context of each learner essay.[46] The native informant was asked to read the Bulgarian and German learner essays carefully and judge all verb phrases in the learner corpora for their grammaticality, indicating the erroneous verb phrases

---

[46] The selection of an American English speaker as a native informant was guided by purely practical considerations of availability – possible bias concerning the native informant's intuitive perception for "correct" use of the progressive and the perfect and the (possible) target norms of use of the progressive and the perfect of the two learner groups will be discussed in more detail in chapters 8 and 9

with a meta-textual error tag[47]. This kind of "problem-oriented annotation" (McEnery et al. 2006: 43) is relevant and useful for a specific research question and does not aim to cover a broad spectrum of linguistic phenomena – in the present study it takes into account only the targetlike and non-targetlike uses of verb phrases in advanced EFL learner writing. Nevertheless, even though the problem-oriented annotation of erroneous verb phrases may not be an exhaustive type of annotation, annotating all of them in two learner corpora comprising some 200,000 tokens each is an enormously time-consuming task which requires of the native informant to read each learner essay carefully.

Since such an error-tagging project would have gone beyond the scope of the present study in the case of the whole *ICLE* components comprising some half a million tokens together and in order to reduce the number of verb phrases to be evaluated and subsequently error-tagged, two sample subcorpora of approximately equal size (almost half the size of each component, c. 100,000 tokens each) of *BUCLE* and *GICLE* were manually extracted on the basis of randomly selected learner essays. The design, number of essays and number of words in the two subcorpora (called for convenience *BUCLE*_110,000 and *GICLE*_110,000) are presented in Table 5.15[48].

Subcorpus	N Words	N Essays
1.*BUCLE*_110,000	112,064	181
2.*GICLE*_110,000	113,230	241
Total	**225,294**	**422**

Table 5.15. Subcorpora design

The problem-oriented annotation procedure for the erroneous verb phrases in the subcorpora followed closely the most recent corpus-based approach to learner use – the computer-aided error analysis (CEA) approach developed by Dagneaux et al. (1998) at the University of Louvain-la-Neuve. CEA is an approach which has originated from traditional Error Analysis, but which "makes full use of advances in C[omputer] L[earner] C[orpus] research" (Dagneaux et al. 1998: 165). Whereas traditional Error Analysis (EA) of the 1970s was mainly used to collect, classify and describe individual learner errors in a somewhat sporadic and anecdotal manner, CEA is a systematic approach which is based on corpora as a

---

[47] The native informant was not informed that the progressive and the perfect verb phrases were at the focus of investigation

[48] The two subcorpora are largely comparable, although the differences and restrictions on the comparability of the design of *BUCLE* and *GICLE* like the length of the essays, the L2 exposure of the learners or the use of reference tools mentioned in 5.2 still apply. The exact corpus make-up and essay codes in the learner subcorpora are given in the appendix.

new source of data and no longer on the individual learner. Considering the weaknesses of traditional EA such as a reliance on fuzzy error categories and the analysis of decontextualised examples in isolation, as well as a total neglect for learner non-use or avoidance of linguistic items, CEA "has inherited the methods, tools and overall rigour of corpus linguistics" (Dagneaux et al. 1998: 173), since it classifies the various error types using predetermined error codes (grammatical errors, lexicogrammatical errors etc.), counts and sorts them in their larger context alongside learner instances of non-use.

Following the CEA procedure of inserting the correct form alongside the erroneous learner form and a searchable meta-textual error tag[49], the native informant was asked to detect all instances of erroneous use of the verb phrases and insert the correct form she would have used in brackets, leaving the original learner form unaffected. The native informant was instructed to label all erroneous instances with a general invariant code 'VT' (for verb tense), which could then easily be searched and sorted by a standard corpus-linguistic software programme like *WordSmith* 4 and further analysed and subclassified into error subtypes concerning the use of tense, aspect or other categories of the verb phrase (carried out by the author of the present work). Subsequently, *WordSmith* concordances were run through all the error tags and the error-tag concordance lines were exported into Excel spreadsheets, where they were further counted, categorised and analysed. A sample concordance of the erroneous verb phrases in *GICLE_110,000* is illustrated in figure 5.7:

Figure 5.7. A sample error-tag concordance based on *GICLE*

The detection of the erroneous verb phrases was largely based on Lennon's definition of an error – "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts" (Lennon 1991:182). Lennon's general definition takes into account both overt formal errors and the so-called "middle ground of advanced learner

---

[49] The annotation followed a simple flat annotation system with error tags integrated in the text

performance" (Lennon 1991: 183) which does not refer to locally-produced formal misformations of e.g. tense-aspect forms, but which depends on the larger linguistic context of the learner production. This middle ground has been variably termed as "dispreferred forms" (Ellis and Barkhuizen 2005: 59), "infelicities" (Granger 2002: 14), or "in-between forms" (Eriksson 2008: 109) and invariably refers to those forms which are not necessarily completely erroneous, but which are still not fully native-like from a native-speaker perspective (cf. Lennon 1991: 184). In his comprehensive study on tense-aspect errors in the Swedish component of *ICLE*, Eriksson notes that:

> there is a tendency of a widened error concept, which does not only cover what is correct and incorrect, but also contains an in-between category which covers use which is not necessarily erroneous, but which is nevertheless not obviously nativelike. (Eriksson 2008: 109)

Such middle-ground or in-between errors will be discussed and analysed in detail in chapter 8 which focuses on the various types of learner misuse of aspect along with a detailed discussion of the native-speaker norm and the problems of error identification, error categorisation and error explanation. The classification of learner use of aspect in the present study thus follows Eriksson's revised model of learner use, which includes both learner native-like use and learner misuse (formal and functional) as defined by Granger (2002: 14), see table 5.16.

Native-like use	Misuse	
	Middle ground (infelicities)	Erroneous use

Table 5.16. Revised model of learner use (adapted from Eriksson 2008: 117)

Since all verb phrases in two POS-tagged subcorpora were error-annotated by the native informant, the problem-oriented annotation approach was combined with a computer-aided error analysis to account for the instances of correct use, misuse or overgeneralisation of aspect forms, as well as undergeneralisation or non-use of aspect forms in the learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000. This combination of a problem-oriented annotation and computer-aided error analysis approach allows for a more in-depth analysis of "what learners get right as well as what they get wrong" (Ellis and Barkhuizen 2005: 70), both from a qualitative and from a quantitative perspective. Granger observes that CEA "does not focus exclusively on errors […] [but] is fully compatible with 'obligatory occasion analysis'" (Granger 2009: 23) and can thus help to determine in a POS-tagged corpus e.g. the number of erroneous auxiliaries out of all auxiliaries.

A similar procedure has been employed in the present study in order to quantify the instances of learner misuse of aspect forms – Pica's (1983) adapted version of obligatory occasion analysis (Brown 1973; Dulay and Burt 1974) – the target-language use analysis quantification method (TLU). TLU was originally developed as a method of morpheme quantification which accounts for both correct learner use and learner non-use of a particular grammatical morpheme (e.g. the *–ing* morpheme) in required, as well as in non-required contexts. The method was proposed in response to the drawbacks of the obligatory occasion analysis which failed to account for the overgeneralisation of morphemes in inappropriate contexts and only analysed the correct or incorrect suppliance of morphemes in appropriate contexts (cf. Pica 1983: 70 – 71). The TLU score for a particular morpheme is calculated by means of the following formula:

TLU = $\dfrac{\text{correct suppliance in obligatory contexts}}{\text{N obligatory contexts + number suppliance in non-obligatory contexts}}$

The TLU score is thus a "ratio which includes in its denominator the sum of both the number of obligatory contexts for suppliance of the morpheme and the number of non-obligatory contexts in which the morpheme is supplied inappropriately" (Pica 1983: 71). Since the TLU measurement was traditionally used to count the number of misformations of morphemes in required contexts and since such overt misformations (e.g. misformations of the past participle form in perfect constructions) were found to be fairly rare in the two learner corpora in the present study representing advanced EFL learners' writing (see chapter 8), the TLU formula was adapted in order to account for the appropriate discourse- and functionally-determined uses of the progressive and the perfect in required contexts, as well as for the inappropriate uses of the progressive and the perfect in non-required contexts. The required contexts in the two learner subcorpora *BUCLE_110,000* and *GICLE_110,000* were determined on the basis of the learner frequency of the progressive and the perfect in the subcorpora in total [50] by subtracting the number of inappropriate uses of the progressive and the perfect in non-required contexts (determined by the error-tagging procedure and extracted and sorted by error-tag concordances) from this frequency and adding the number of non-uses of the progressive and the perfect in required progressive and perfect contexts. In this way,

---

[50] The total learner frequencies for the progressive and the perfect in the learner subcorpora, as well as the frequency of the finite verb phrases were calculated by the same quantitative procedures outlined in section 5.4.

two types of measurements were used to quantify learner misuse of the progressive and the perfect: learner non-use or undergeneralisation of the progressive and the perfect in required progressive and perfect contexts, and learner overgeneralisation of the progressive and the perfect in the contexts of other tense-aspect forms. To illustrate, the percentage of non-use of e.g. the perfect in required present and past perfect contexts was calculated by dividing the number of perfect forms which were not supplied in required perfect contexts by the number of all required contexts for the perfect[51]:

$$\% \text{ Non-use }_{\text{perfect aspect}} = \frac{\text{N non-use in required perfect contexts}}{\text{N required perfect contexts}} \times 100$$

The second measurement calculates the proportions of the progressive and perfect forms used incorrectly in non-progressive and non-perfect contexts (i.e. overgeneralisation) in the learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000 by means of the following formula:

$$\% \text{ Over-suppliance }_{\text{progressive/perfect}} = \frac{\text{N over-suppliance }_{\text{progressive/perfect}}}{\text{N finite verb phrases} - \text{N required contexts }_{\text{progressive/perfect}}} \times 100$$

The second formula was used in order to account for the percentage of "encroachment" of progressive and perfect verb forms on the required contexts of other non-progressive, non-perfect forms in the two learner subcorpora. The application of these two measurements as an adapted form of the TLU morpheme quantification method proposed by Pica (1983) to error-annotated and POS-tagged learner corpora provides for valuable insights into learners' distribution patterns of the progressive and the perfect, as well as learners' misuse rates in terms of two types of misuse: undergeneralisation or non-use of aspect forms in required contexts and their incorrect overgeneralisation to non-required contexts. Applied to advanced EFL learners' writing, these measurements highlight the problematic areas with regard to learners' form-function mapping of aspect in English and can be employed in addition to the methods measuring learner over- and underuse traditionally employed within the Contrastive Interlanguage Analysis framework (cf. Granger 1996). Lastly, they serve as a good starting point for qualitative examinations of the major areas of functional confusion between tense-aspect forms in advanced EFL learner writing.

---

[51] The same formula was applied for the non-use of the progressive in required progressive contexts

The present chapter introduced the corpus design, software tools and research methodology employed in the analysis of aspect use by advanced Bulgarian and German EFL learners of English. The following chapters 6, 7 and 8 will present the quantitative and qualitative results obtained on the basis of the corpus data and methodology outlined above.

6. Aspect Use in Learner and Native Writing: Quantitative Results

On the basis of the corpus data and the methodology described in the previous chapter, the present chapter deals with the description of the quantitative differences between learner and native use of the progressive and the perfect. The main purpose of this chapter is to illustrate and explain the overall distribution patterns of aspect forms in learner and native writing and to detect learner deviations from the native-speaker corpus-based norm such as underuse or overuse of the progressive and the perfect. The chapter will start with an outline of the results for the overall frequency distribution of the finite verb phrases in the two learner and four native corpora under scrutiny and will proceed with a contrastive comparison between the individual frequencies of use of the progressive and the perfect in learner and native writing. Finally, a brief summary of the quantitative results and a comparison with the previous frequency results for the progressive and the perfect in learner writing reviewed in chapter 2 will be presented.

6.1. Frequency Distribution of the Finite Verb Phrases in Learner and Native Writing

Calculating the distribution of finite forms in a POS-tagged corpus has two major advantages: on the one hand, it can be used as a benchmark for comparison between different text types or registers in order to establish differences in terms of the degrees of conceptual orality of written texts (cf. Koch and Oesterreicher 1985), and on the other, it can be used as the basis for the second type of measurement of aspect forms described in chapter 2 (Smitterberg's (2005: 44) adapted V-coefficient measurement) alongside the normalised frequency per 1,000 words. Table 6.1. illustrates the number of finite verb tags and the total number of POS tags, whereas figure X graphically represents the proportions of finite verb tags of all tags in the six corpora.

	BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB_F*	*FROWN_F*
finite verb POS tags	22377	24571	8068	16457	8567	9721
total POS tags	189934	214954	74627	142020	104250	114948

Table 6.1. Comparison of the finite verb POS tags and all POS tags across the six corpora
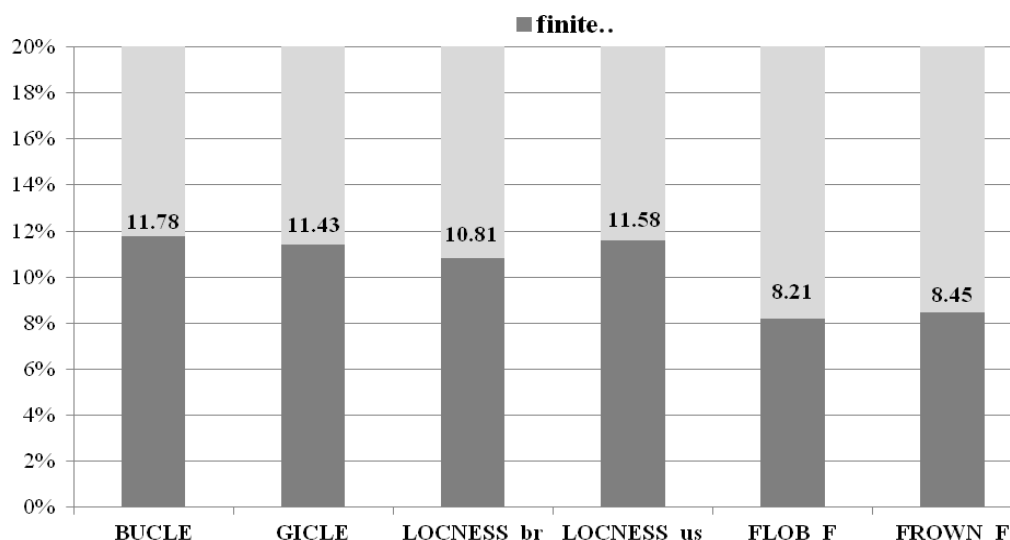
Figure 6.1. Ratio of the finite verb tags of all POS tags in the six corpora

An initial comparison between the frequencies and proportions of finite verb tags of all tags in the six corpora already reveals interesting insights about the degrees of 'verbiness' in learner, novice native and expert native writing. To illustrate, the Log Likelihood ratio test shows highly significant differences between learner writing and expert native writing (p <0.001) and likewise significant differences between native novice and native expert writing (p<0.001). The Bulgarian learner corpus *BUCLE* is the 'verbiest' of all six corpora, being significantly verbier than the German learner corpus *GICLE* (p<0.01), whereas the British expert writing corpus *FLOB*_F is the least verby or the 'nouniest' corpus of all six corpora.[52] There are no significant differences between the frequencies of use of the finite verb tags in the American part of the *LOCNESS* corpus *LOCNESS*_us and either learner corpus; however, there are significant differences between the novice native corpora and the expert native corpora (p<0.001). Thus, the two learner and four native corpora can be placed on the following scale, ranging from 'verby' to 'nouny' corpora[53]:

BUCLE > LOCNESS_us > GICLE > LOCNESS_br > FROWN_F > FLOB_F

Three major factors need to be taken into account for the interpretation of these preliminary results: Bulgarian and German EFL learners' writing style in English, the writing competence of native speakers of English, as well as issues of data sampling in the learner and

[52] Notably, the differences measured in percent are not very big (only about 3.5 %); however, they still indicate possible differences in 'verbiness' between the six corpora
[53] Multiple '>' signs correspond to greater differences

native corpora. The finding that Bulgarian and German EFL learners overuse finite verb phrases in comparison to expert native writers is not surprising insofar as it confirms previous results with regard to learners' higher use of verbs at the expense of nouns and their more verbal style of writing (cf. Eriksson 2008: 81). Eriksson (2008) and Altenberg (1997) note that Swedish EFL learners use considerably more finite verb phrases than native speakers of English – a finding which Eriksson attributes to the L1-L2 differences between Swedish and English (Eriksson 2008: 81), but which in view of the 'verbiness' scale illustrated above calls for a different explanation. This finding ties in with Gilquin and Paquot's (2007: 7 – 9) observation that advanced learners of English from various mother tongue backgrounds use a number of colloquial features in their written production which are more typical of speech than of academic writing and which contribute to the greater colloquial overtone of learner texts in comparison to native-speaker texts. Gilquin and Paquot's (2007: 9) conclusion that both EFL learners and native speakers with little or no expertise in writing resort to spoken features in academic writing can be confirmed on the basis of the verb tags' distribution in the learner and native corpora under scrutiny: both the learner corpora and the novice native corpora *LOCNESS*_br and in particular *LOCNESS*_us feature a similar overuse of finite verb phrases in comparison to the expert native corpora *FLOB*_F and *FROWN*_F.

Thus, both the EFL learners and the native student writers in the present sample move along a formality-colloquiality continuum which reflects their writing competence in expository writing and which at the same time reflects the orality-literacy continuum proposed by Koch and Oesterreicher (1985). One striking difference that needs a further comment is the fact that *BUCLE* is not only the verbiest corpus of all six corpora, but also somewhat verbier than *GICLE*. Two explanations appear plausible here: learners' proficiency and the rhetorical and organisational patterns of written texts in German and Bulgarian. With respect to the former factor, since the majority of the *GICLE* learners have had greater L2 exposure to English in an English-speaking environment and together with native-speaker teachers of English in the German EFL classroom (cf. Granger et al. 2002; 2009; Lorenz 2002: 102 in Granger et al. 2002), and since the *GICLE* learners have been rated as more advanced learners on average than *BUCLE* learners, it can be safely assumed that the *GICLE* learners are more proficient users of English on the whole, both in speech and in writing. With respect to the differences in the organisation of academic texts, we can expect of German EFL learners' texts to reflect the so-called Teutonic academic style (as would be the case in original German texts), i.e. academic style which relies on "a large number of

nominalizations, overloaded phrases and agentless passives" (Connor 1996: 54). Given that *GICLE* learners have already had instruction in German academic writing in a university context, we may assume that their reliance on nominalisations may be transferred to L2 English to the effect that the high frequency of verbs is slightly reduced; nevertheless, *GICLE* learners still use less finite verbs than *BUCLE* learners.

Lastly, a major factor which influences the distribution of finite verb phrases across the six corpora concerns the sampling of the learner and native corpora. Even though the learner and novice native control corpora in the present study have been sampled according to strict and explicit criteria in order to match learner proficiency and native-speaker writing competence, differences between the text types in the corpora are bound to occur. These differences reflect not only individual learner differences, but also the wide range of essay prompts, especially in *GICLE* and in *LOCNESS*, many of which vague and prompting the students to narrate personal stories that involve a more personal and less abstract style. To illustrate, the verbiest and closest to the learner corpora native control corpus *LOCNESS*_us features over fifty different essay topics, many of which fairly general: "Homosexuality", "Sex in Schools", "Abortion", "Gender roles" etc. (cf. Granger et al. 2002; 2009). These topics often encourage native students with little or no expertise in argumentative writing to overuse personal finite stance markers like *I think* and *I believe* (cf. Eriksson 2008: 81), as in the following example:

> 6.1. **I mean** if the model in the commercial can look like that because she uses that certain product -- so **can I** (yeah right.) **I believe** that females are given a false sense of hope or expectation, because **I know** that how ever much makeup **I put** on, **I will** never look like Christie Brinkley […].<*ICLE*-US-SCU-0004.2>

Still, the verbiness cline illustrated above implies that the two most important factors determining the frequency of use of finite verb phrases in expository writing are the writing competence of native and learner writers, together with the learner proficiency. The overall learner and native frequencies of finite verb phrases are particularly important for the comparison of aspect forms across learner and native writing, since they serve as the basis for Smitterberg's (2005: 44) adapted V-coefficient measurement alongside the normalised frequencies per 1,000 words. The quantitative comparison of the progressive and the perfect verb forms in the learner and native corpora will be dealt with in the next section.

6.2. Frequency Distribution of the Progressive in Learner and Native Writing

Two types of measurement were used to compare the frequencies of use of the progressive in the learner and native corpora: a normalised frequency per 1,000 words and Smitterberg's (2005: 44) adapted V-coefficient measuring the relative frequency of the progressive verb phrases in relation to the finite verb phrases. Table 6.2. illustrates the first measurement: the absolute and normalised frequencies of the progressive (per 1,000 words) in the six corpora, whereas figure 6.3. shows a comparison of the normalised frequencies with the frequencies of use of the progressive identified by the LGSWE for academic writing, news and fiction (Biber et al. 1999: 461).

	BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB_F*	*FROWN_F*
progressives	385	603	200	568	256	201
progressives per 1,000 words	1.9	2.7	2.5	3.8	2.9	2.1
total number of words	199,249	226,503	79,228	149,573	88,574	96,587

Table 6.2. Absolute and normalised frequencies of the progressive in the six corpora
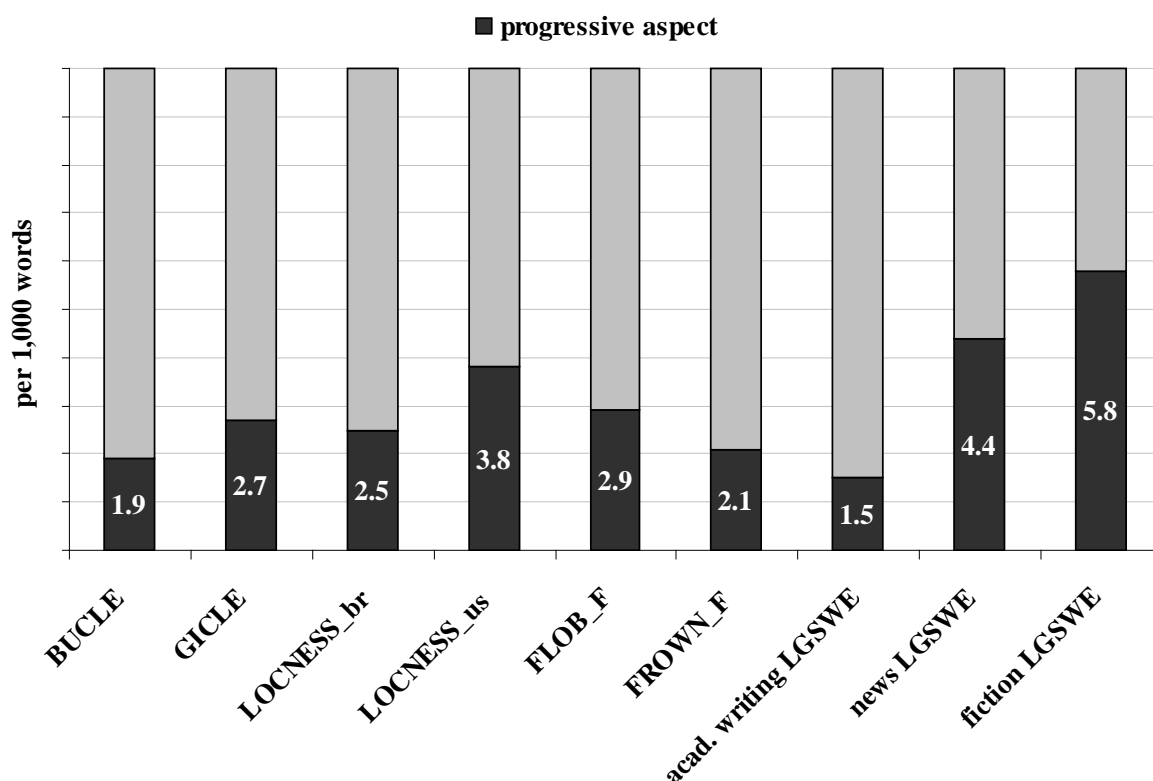
Figure 6.2. Comparison of the normalised frequencies of the progressive in the learner and native corpora with the LGSWE frequencies for academic writing, fiction and news (cf. Biber et al. 1999: 461)

The Bulgarian learner corpus features the lowest normalised frequency of all six corpora (1.9 instances per 1,000 words), whereas the American novice writer corpus the highest (3.8 instances per 1,000 words). The frequencies of use of the progressive in all six corpora lie within the LGSWE academic writing-news range, but are in general significantly higher than the LGSWE frequency for academic writing (with the exception of *BUCLE*). Furthermore, there are statistically significant differences between the learner and native corpora: starting with the native-speaker frequencies, the novice native corpora confirm the expectations concerning a higher preference for the progressive in American English, since *LOCNESS*_us features significantly more progressives than *LOCNESS*_br (p<0.001). Contrary to the expectations, however, are the frequencies of use of the progressive in the expert native corpora: the British corpus *FLOB*_F which features more progressives than the American corpus *FROWN*_F (p<0.001). Focussing on the learner frequencies of use, the Bulgarian learner corpus exhibits significant underuse of the progressive in comparison to all other corpora (p<0.001) with the exception of *FROWN*_F; in contrast, the German learner corpus exhibits much less significant differences to the native control corpora. In comparison to *LOCNESS*_us, German EFL learners underuse the progressive significantly (p<0.001), whereas in comparison to *FROWN*_F they overuse it (p<0.01). There are no significant differences between the German learner values and either British value for the progressive.

The second measurement of progressive forms is Smitterberg's adapted V-coefficient measurement which involves a more refined normalisation procedure of aspect forms. It takes into consideration the number of finite verb phrases as a basis for comparison in order to neutralise the effects of the verbiness cline presented in the previous section and calculates the frequencies of use of the progressive in the six corpora as a proportion of all finite verb phrases. The proportions in % are graphically represented in figure 6.3.
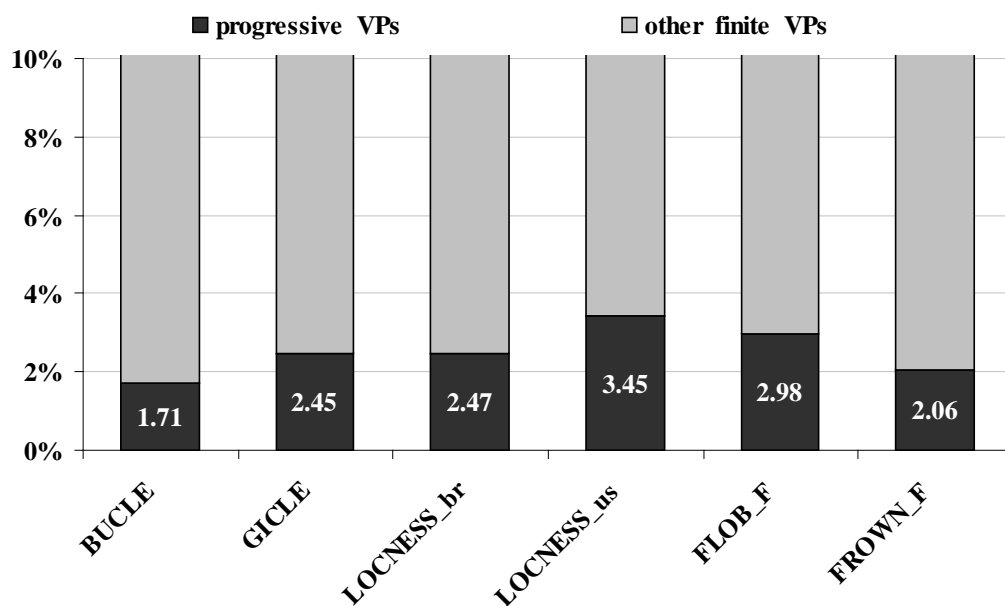
Figure 6.3. Progressive verb phrases proportionate to the number of finite verb phrases in %

The frequencies of use of the progressive proportionate to the number of finite verb phrases in the six corpora reveal a similar, although a more refined picture with regard to the quantitative comparison between learner and native writing: considering the number of finite verb phrases, the differences between the learner and native samples can be better neutralised and are thus a more suitable input to statistical tests measuring variation (cf. Smitterberg 2005: 44). Indeed, the differences between the learner and native proportions of progressive verb phrases are somewhat more significant in comparison to the normalised frequencies of use presented above: thus, *BUCLE* exhibits a significant underuse of the progressive in comparison to all other corpora (p<0.001), including *FROWN*_F (p<0.05), whereas *GICLE* also shows significant underuse in comparison to all other control corpora (p<0.05) with the exception of *LOCNESS*_br. The differences between the native corpora remain the same, with *FLOB*_F featuring – contrary to the expectations – surprisingly high number of progressive verb phrases in comparison to *FROWN*_F, as well as to the novice native British corpus *LOCNESS*_br.

Concerning the temporal specification of the progressives in the learner and native corpora, the present progressive is the preferred option and the dominant progressive form for all six corpora (accounting for 2/3 of all progressive uses), followed by the past progressive

and a minor use of the combination of aspect forms – the perfect progressive (present and past) (see figure 6.4).
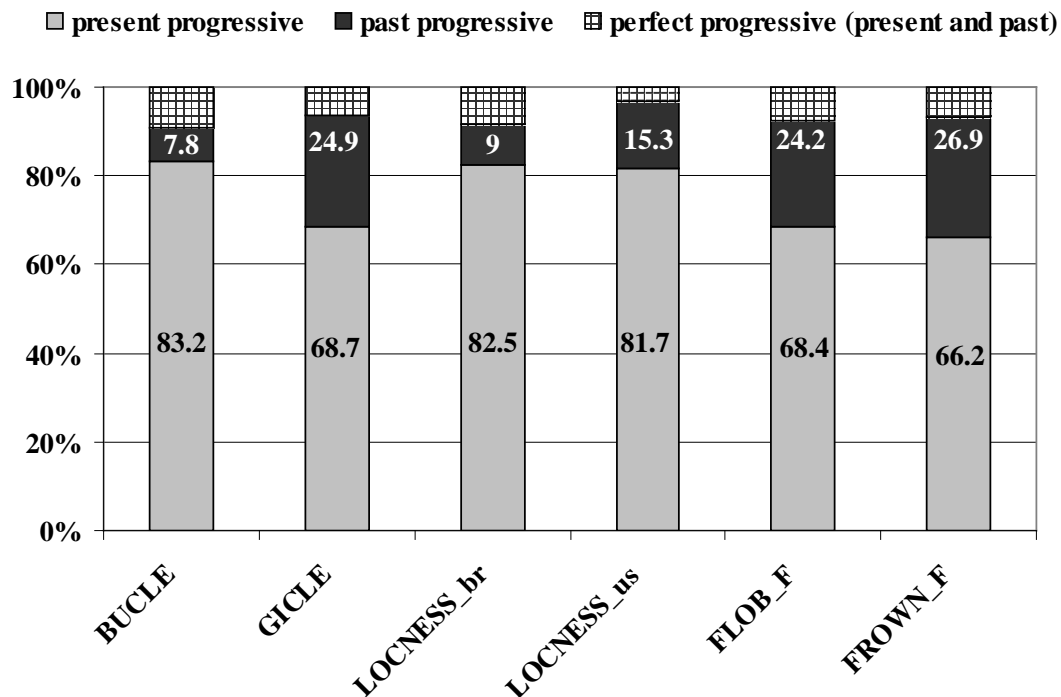
Figure 6.4. Temporal specification of the progressive and combination with the perfect in the learner and native corpora

Some interesting trends emerge here: *GICLE* and the two expert native corpora *FLOB*_F and *FROWN*_F show similarly high rates for the past progressive which are significantly higher than the proportions of the past progressive in the other three corpora *BUCLE*, *LOCNESS*_br and *LOCNESS*_us (p<0.05), and which at the same time presuppose an altogether higher use of past forms in these corpora. The differences in the use of the past progressive and past forms in general between the corpora can be explained with the differences between the internal make-up of the corpora on the one hand (and in particular the design of the expert native corpora), and with learners' lack of register awareness for the text type 'argumentative essay' on the other. The former criterion becomes obvious when looking into the individual texts of *FLOB*_F and *FROWN*_F, many of which in addition to their commentaries on current affairs in Britain and in the US often refer to past events using the simple past, as in the following example:

6.2. F01 Certainly, what **happened** between the late fifties and the early seventies **was** not a political revolution, not a revolution in economic thought and practice; but it **was**, I believe, a transformation in the opportunities and freedoms available both to the

majority as a whole and to distinctive individuals and groups within that majority. <*FLOB*_F, F01 27 – F01 33>

Indeed, a brief look at the overall distribution of tense-aspect forms in the six corpora (see figure 6.5) confirms this observation by showing that simple past forms are similarly frequent in *GICLE*, *FLOB*_F and *FROWN*_F (18%, 27.6% and 26.5% respectively), whereas *BUCLE*, *LOCNESS*_br and *LOCNESS*_us feature much lower proportions of simple past verb phrases (7.7%, 9.9% and 16% respectively). Therefore, the higher frequencies of the past progressive which co-occur with the simple past in these three corpora are not at all surprising.
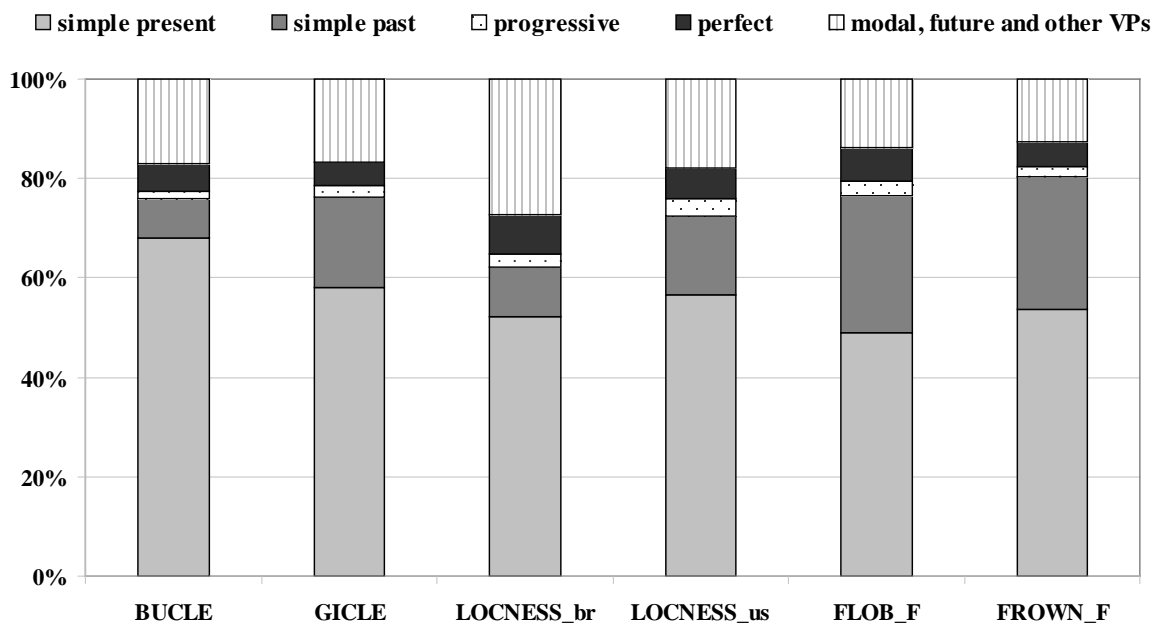
Figure 6.5. General distribution of tense-aspect forms in the six corpora

The second reason for the higher use of simple past and past progressive forms in *GICLE* may be explained with EFL learners' inability to produceabstract arguments away from straightforward narration (following Biber's multidimensional model – dimension 2, for which the use of past tense verbs counts as a positive feature in favour of narration vs. non-narration, see also Biber 1988; Conrad and Biber 2001) . Notably, this inability is also triggered by some of the essay prompts in the learner and native samples, since many of the topics in *GICLE* are as vague as the topics in *LOCNESS*_us and thus often encourage learners to narrate personal stories that happened in the past (e.g. 'Do it yourself' or 'The pleasures of cycling', 'Someone I admire' cf. Granger et al. 2002; 2009). Consequently, learners and novice native writers start relating personal accounts in which they resort to the use of the simple past and the past progressive, as illustrated in the following example:

6.3. The first thing that **caught** my eye when I **met** this extremely handsome, young man **was** the captivating, warm expression in his sparkling, deep ocean-blue eyes as they **peered** straight at me. Strolling towards me, his thick, shiny, golden hair **glowed** in the sunshine and a light, refreshing summer breeze **played** with his stubborn curls. <*ICLE*-GE-AUG-0002.1>

Having considered the overall distribution of finite verb phrases and progressive verb forms in the learner and native corpora, two important trends need to be summarised and considered for a more detailed discussion of the results: the overuse of finite verb forms and the underuse of progressive verb forms by both Bulgarian and German EFL learners. An attempt at explaining the overuse of finite verb forms by both learner groups was made earlier: irrespective of their mother-tongue background, both Bulgarian and German EFL learners overuse colloquial markers in expository writing in contrast to professional native English writers (cf. Gilquin and Paquot 2007), mostly due to their lack of writing expertise in English. An explanation of the underuse of the progressive aspect by Bulgarian and German EFL learners requires more thought: although both Bulgarian and German lack a progressive proper, the correlation between underuse and the native-language influence is not as straightforward as it seems. Even though L1 Bulgarian features two institutionalised morphosyntactic forms which relate meanings of progressivity (imperfective verb forms and the past imperfect), German EFL learners use more progressives than Bulgarian EFL learners in English, and both learner groups underuse the progressive in comparison with the native corpora. These particular findings and a possible explanation of the differences in the light of L1-transfer will be dealt with in chapter 9. Before addressing them in more detail, the overall distribution of perfect verb forms in the learner and native corpora will be addressed first.

6.3. Frequency Distribution of the Perfect in Learner and Native Writing

The same two measures used for the comparison of the frequencies of use of the progressive in the learner and native corpora were used for the comparison of perfect verb forms. Table 6.3 illustrates the absolute and normalised frequencies (per 1,000 words) of the perfect in the six corpora, whereas figure 6.6 shows a comparison of the normalised frequencies with the frequencies of use of the perfect outlined by the LGSWE for academic writing, news and fiction (Biber et al. 1999: 461).

	BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB*_F	*FROWN*_F
perfect	1200	1131	655	998	581	504
perfect per 1,000 words	6	4.9	8.3	6.7	6.5	5.2
total number of words	199,249	226,503	79,228	149,573	88,574	96,587

Table 6.3. Absolute and normalised frequencies of the perfect in the six corpora
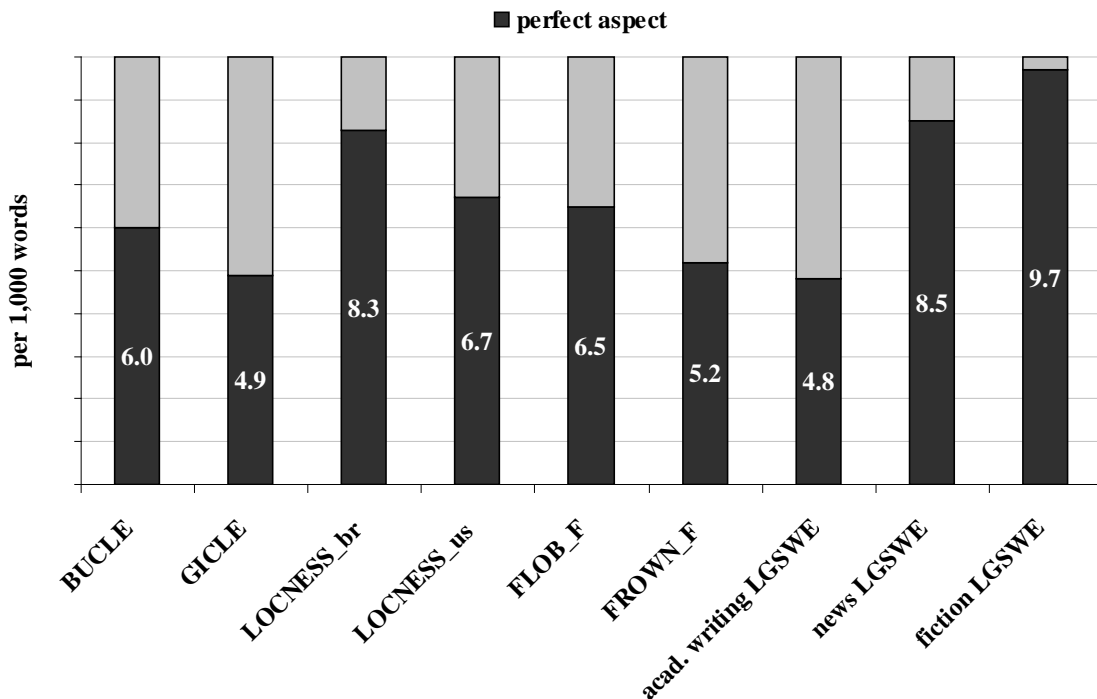
Figure 6.6. Comparison of the normalised frequencies of the perfect in the learner and native corpora with the LGSWE frequencies for academic writing, fiction and news (cf. Biber et al. 1999: 461)

The German learner corpus *GICLE* features the lowest normalised frequency of use of the perfect (4.9 instances in 1,000 words) of all six corpora, whereas the British novice writer corpus *LOCNESS*_br the highest (8.3 instances in 1,000 words); similar to the frequencies of use of the progressive, all six corpora lie within the LGSWE academic writing-news range. Again, there are statistically significant differences between the native corpora: unsurprisingly, both the novice and the expert British control corpora show a significantly higher preference for the perfect than the American novice and expert corpora ($p<0.01$). Not as uniform are the differences between the learner and native corpora: thus, *BUCLE* exhibits a significant underuse of the perfect in comparison with both novice native corpora *LOCNESS*_us and *LOCNESS*_br ($p<0.05$), but at the same time, it features a significant overuse of the perfect in comparison with *FROWN*_F ($p<0.01$) and no difference to *FLOB*_F. *GICLE* learners' normalised frequencies are more straightforward insofar as they are significantly lower than the frequencies of use of the perfect in all other corpora ($p<0.001$)

except for *FROWN*_F (no significant difference). The adapted V-coefficient measurement for perfect forms presented in proportion to all finite verb forms is graphically represented in figure 6.7.
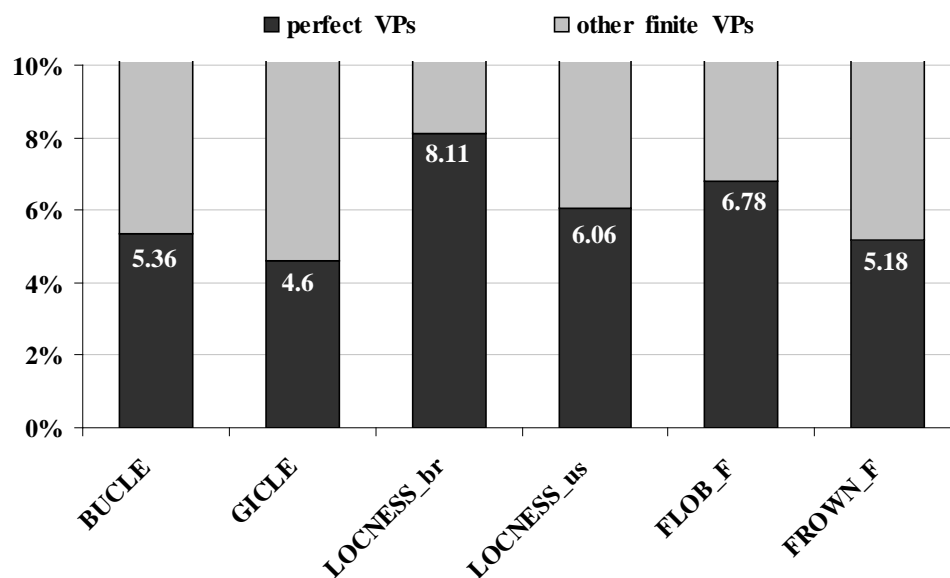
Figure 6.7. Perfect verb phrases proportionate to the number of finite verb phrases in %

Like with the progressive proportions, the frequencies of use of the perfect proportionate to the number of finite verb phrases again reveal a somewhat different picture: whereas *BUCLE* learners use more perfects than *GICLE* learners ($p<0.001$), they still underuse them significantly in comparison to all native control corpora ($p<0.01$) with the exception of *FROWN*_F. Likewise, *GICLE* learners underuse the perfect in comparison to all native control corpora highly significantly ($p<0.01$). Concerning the temporal specification of the perfect in the learner and native corpora, the present perfect is again the preferred option and the dominant perfect form in all six corpora, followed by the past perfect.
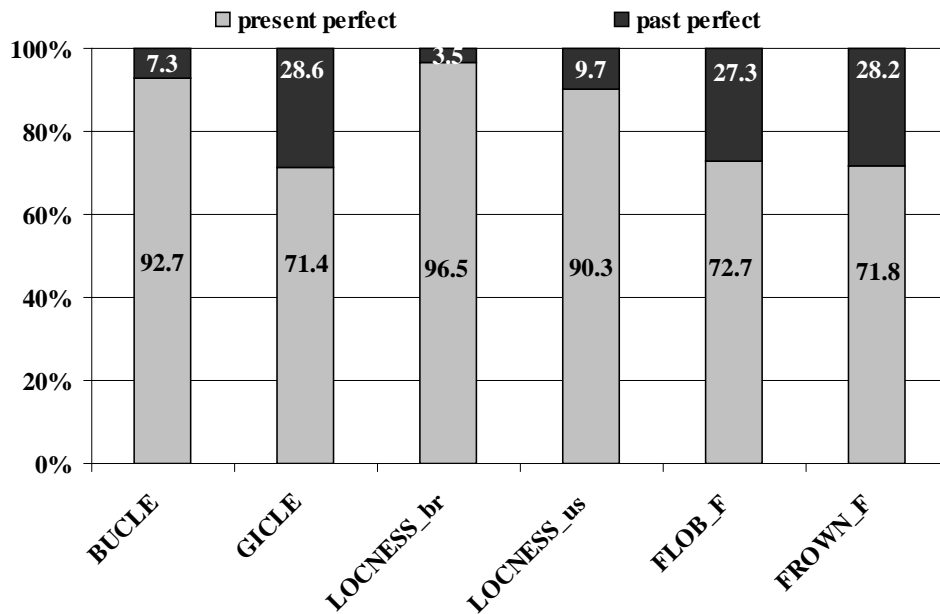
Figure 6.8. Temporal specification of the perfect in the learner and native corpora

The temporal specification of perfect verb forms in the six corpora reveals the same striking similarities between the corpora as with the past progressive: on the one hand, there are similarities between *GICLE* and the two expert native corpora *FLOB*_F and *FROWN*_F, and on the other, between *BUCLE*, *LOCNESS*_br and *LOCNESS*_us. Whereas the German learner corpus and the expert native corpora feature a higher use of the past perfect which accounts for over a quarter of all perfect uses, the Bulgarian learner corpus and the novice native corpora overwhelmingly prefer the present perfect (over 92% of all uses). The similarities in the use of past perfect forms in *GICLE*, *FLOB*_F and *FROWN*_F deserve an explanation: apart from the higher proportions of narration in the learner and native corpora presupposing higher proportions of past forms altogether, a higher use of conditional clauses may also be the reason for the increased use of the past perfect. However, a closer look at the learner data in *GICLE* reveals that only about 7% of all past perfect uses occur in conditional clauses; still, the vast majority of the past perfect uses in *GICLE* occur in relative or adverbial subordinate clauses (over 70% of all past perfect uses), as illustrated in examples 6.4 and 6.5[54]:

> 6.4. Finally as **I had decided** to leave my bed, to stand up in order to take off the receiver, I heard the well-known voice of my mother […] <*ICLE*-GE-AUG-0024.1>

---

[54] A more detailed investigation of the distribution of perfect forms in main and subordinate clauses will be carried out in the next chapter.

6.5. After five weeks both were allowed to leave the hospital and as **they had learned** to love each other, they decided to get married. <*ICLE*-GE-SAL-0023.5>

In these two examples, both subordinate clauses are introduced by the subordinating conjunction 'as', which is used to introduce an instantaneous event like the decision to get out of bed as in example 6.4, or the more durative act of getting to love each other, as illustrated in example 6.5. Here, the idea of lexical transfer from the German conjunction *als*, which is used to introduce subordinate clauses specifying a temporal circumstance, is not far-fetched: the direct translations of example 6.4 into German *als ich mich entschied* and example 6.5 into *als sie sich lieben lernten* seem perfectly acceptable and plausible in the context of these two sentences. In addition, there are a number of uses of the past perfect in main clauses, where the past perfect is used to relate actions or events happening as part of the main story line:

6.6. A couple of friends and I **had been going out** together and Wolfgang **had been** driving the car. Unluckily I **had forgotten** my scarf in his car and so I had to call him. <*ICLE*-GE-AUG-0023.1>

6.7. Yesterday I **had been to** the formal ball the mayor of our town had organized to collect some money for Norogachic, a very poor village deep in the dry and bare Mexican Sierra Madre Occidental. <*ICLE*-GE-AUG-0061.1>

In examples 6.6. and 6.7, the past perfect is used as the main narrative tense relating events and actions that happened in the past; in both cases, the simple past would have been the preferred option in the target; however, the German learners have opted out for the more complex past perfect form instead of the simple past[55]. These examples suggest a possible tendency for *GICLE* learners to overgeneralise the past perfect to non-past-perfect contexts to a much greater extent than *BUCLE* learners: on the basis of the learner frequencies and the few examples presented above, this assumption needs further clarification and will be addressed in more detail in the chapters 8 and 9. The next section will offer a summary of the quantitative findings and a brief comparison with the results obtained in previous learner corpus studies on the use of aspect in writing.

---

[55] A detailed analysis of learner misuse of the progressive and the perfect will be presented in chapter 8.

6.4. Summary

The quantitative results for the use of the progressive and perfect forms in advanced Bulgarian and German EFL learners' written English are surprising in the sense that the two learner groups deviate from the corpus-based target norm in a somewhat different manner than expected and proposed by the findings of previous learner corpus studies (see section 4.3). German EFL learners use significantly more progressives than Bulgarian EFL learners; however, they do not overuse them (as previously reported) in comparison to the majority of the native-speaker English corpora used as benchmarks in the present study, but rather, underuse them significantly. The adapted V-coefficient measurement developed specifically for the present study to compare the progressive ratios across the learner and native corpora in comparison to all finite verb forms also corroborates this finding.

The results for the perfect in learner writing are also unexpected: although Bulgarian EFL learners use significantly more perfect forms than German EFL learners, they still underuse them significantly in comparison to all native corpora with the exception of *FROWN*_F. Likewise, German EFL learners also underuse the perfect highly significantly in comparison to all native corpora without exception. The latter finding is indeed surprising, having in mind the formal similarity between the German *Perfekt* and the English perfect and the probability for learners' consequent functional confusion between the two; at the same time, it confirms Davydova's latest findings with respect to German EFL learners who "[b]eing unsure of the exact meanings conveyed by the English perfect […] try to avoid using this form altogether, replacing it with a semantically simpler form – the preterite" (Davydova 2011: 288).

Finally, there is a significant variation between both the learner and the native corpora – even though certain similarities between e.g. the use of the perfect in the Bulgarian learner corpus and the British corpora or the use of the perfect in the German learner corpus and the American corpora may seem plausible, the native corpora themselves pattern differently, and even contrary to the expectations set by previous corpus-based studies and grammars. Whereas the native novice corpora *LOCNESS*_br and *LOCNESS*_us mirror the distribution of the progressive and the perfect in British and American English identified by previous studies

like Biber et al. (1999) – i.e. more progressives are typical of American English and more perfects of British English (see also chapter 3), the expert corpora *FLOB*_F and *FROWN*_F (subcorpora of *FLOB* and *FROWN* respectively) pattern accordingly only in the case of the perfect aspect; surprisingly, *FLOB*_F features more progressives than its American counterpart *FROWN*_F. In addition to the quantitative analysis of progressive and perfect forms in learner and native writing outlined by the present chapter, the following chapter will deal with a more in-depth distributional analysis of progressive and perfect verb forms in learner and native writing.

7. Lexicogrammatical Variation in the Use of Aspect in Learner and Native Writing

The distributional analysis of the use of the progressive and the perfect in learner writing outlined in the previous chapter is only the first step towards a more comprehensive description of aspect use in advanced learners' English; a second step involves a description of the specific L2-features which characterise aspect use in learner English and which are often located at the interface between lexis, syntax and semantics. The aim of the present chapter is to examine and interpret the results of the distributional analysis in the light of the theoretical frameworks concerning the second-language acquisition and use of English aspect and in particular in the light of the claims of the Aspect and Discourse Hypotheses. Furthermore, learner uses of perfect aspect forms with accompanying temporal adverbials, together with the tendency for EFL learners to employ conversational features like contracted auxiliaries in expository writing will also be examined and compared with novice and expert native writing. The description will first cover a quantitative and a qualitative investigation of learner and native uses of the progressive and the perfect under the influence of the inherent lexical aspect of verbs (sections 7.1 and 7.2.); next, the distribution of progressive and perfect verb phrases across main and subordinate clauses will be studied (section 7.3.). Finally, a description of the co-occurrence of aspect forms with temporal adverbials (section 7.4.) and with contracted auxiliaries (section 7.5.) will be delivered.

7.1. The Progressive and the Influence of Inherent Lexical Aspect

Since lexical aspect is a compositional property of verb phrases as whole entities and includes their nominal arguments such as objects and complements (cf. Haznedar 2007: 391), all progressive verb phrases extracted from the six corpora *BUCLE*, *GICLE*, *LOCNESS*_br, *LOCNESS*_us, *FLOB*_F and *FROWN*_F were carefully examined in their concordance contexts with regard to inherent semantic properties like telicity and punctuality in order to classify them in accordance with Vendler's (1957) four inherent lexical aspect categories (cf. Haznedar 2007; Bardovi-Harlig 2000; 2002; Brinton 2000; Brinton 1998; Collins 2002; Bardovi-Harlig and Reynolds 1995 etc.). The categorisation of all progressive forms into Vendler's fourfold division allows for an across-category analysis of one verbal inflection

(the progressive) in order to compare the percentage of all progressives that are activities, states, accomplishments and achievements (cf. Bardovi-Harlig 2000; 2002). However, it will be first presented in terms of a twofold distinction between telic and atelic progressive VPs (Figure 7.1), before a more fine-grained differentiation between all four inherent lexical categories is carried out (Figure 7.2.).
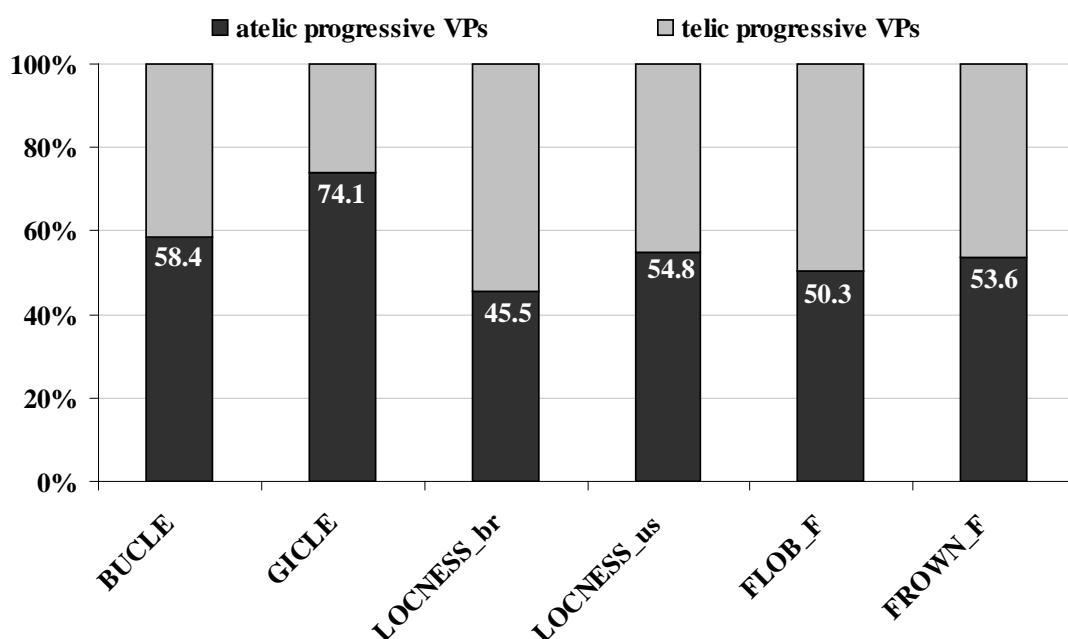
Figure 7.1. Distribution of progressive verb phrases across telic and atelic lexical verbs

The overall distribution of telic and atelic verb phrases marked for the progressive aspect reveals an interesting picture: on average, the novice and expert native corpora feature fewer atelic verbs in the progressive than the two learner corpora, *GICLE* having the highest ratio of atelic verbs in the progressive and *LOCNESS_*br the lowest. This finding already points to the fact that learners prefer atelic verbs – activities or states – with the progressive to a greater extent than native speakers of English. Furthermore, there are statistically significant differences between the two learner groups and the novice and expert native writers in the sample – there are significantly more atelic verbs in the progressive in *GICLE* than in all other corpora ($p<0.001$), whereas *BUCLE* learners overuse atelic verbs only in comparison to the British novice native writers in *LOCNESS_*br ($p<0.05$). There are no such significant differences between the novice and expert native corpora.

An even more interesting picture appears when breaking down the findings into all four of Vendler's (1957) categories – states, activities, accomplishments and achievements (see Figure 7.2). The most striking difference between the learner and native corpora concerns the marking of stative verbs for the progressive – both *GICLE* and *BUCLE* feature considerably more stative verbs in the progressive than the novice and expert native corpora; *GICLE* learners overuse stative verbs in the progressive at the significance level of 0.001 in comparison to all other corpora, whereas *BUCLE* learners overuse stative verbs at the significance level of 0.01 in comparison to both *LOCNESS*_br and *FROWN*_F. In terms of the use of activity verbs in the progressive, both learner corpora feature more activities in the progressive on average; moreover, *GICLE* learners overwhelmingly prefer them ($p < 0.001$) in comparison to all other corpora apart from *FROWN*_F, whereas no such significant differences are apparent between *BUCLE* learners and the native corpora. These findings already suggest that although advanced Bulgarian and German EFL learners prefer to use atelic verbs with the progressive, they do not do this in full accordance with the Aspect Hypothesis, since both learner groups overuse stative verbs in the progressive, thus contradicting the fourth claim stating that "[p]rogressive markings are not incorrectly overextended to stative verbs" (Andersen and Shirai 1996: 533).
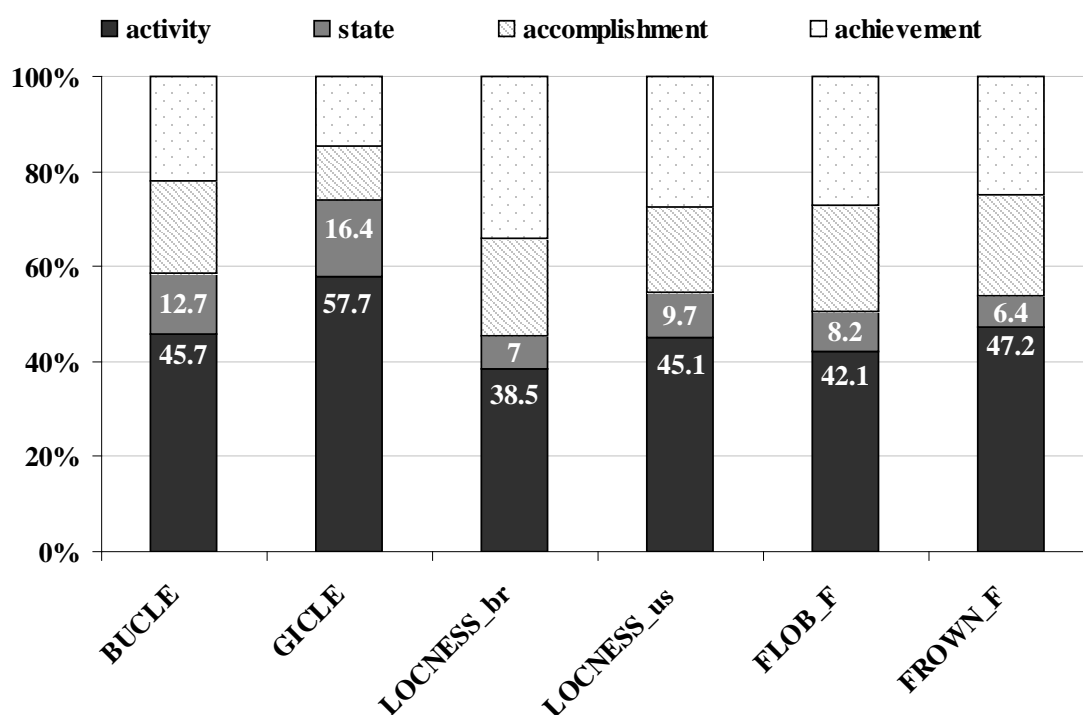
Figure 7.2. Distribution of the progressive verb phrases across all four Vendler types

A closer look at the most frequent 20 verb types in the progressive in the learner and native corpora in comparison to the most frequent 20 verb types in the written part of the BNC (table 7.1) confirms this trend: on average, there are considerably more atelic verbs altogether in the top 20 verb types in the progressive (highlighted in the darker cells), as well as more stative verbs (given in bold) in *BUCLE* and *GICLE* than in the four native corpora and the BNC. Particularly striking is the high number of stative verbs among the top 20 progressive verbs in *GICLE* – there are 7 stative verbs altogether (*sit, be, live, think, stand, watch* and *lie*) in contrast to the only two stative verbs in the progressive *be* and *have* among the top 20 verbs in the progressive in the BNC, which are also present among the top 20 verbs in all four native corpora.

BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB*_F	*FROWN*_F	BNC_written
try	go	**be**	**be**	**be**	do	**be**
be	**sit**	become	try	do	**be**	go
become	talk	take	go	go	try	**have**
live	try	go	become	get	come	use
dream	get	try	do	**have**	say	make
do	look	increase	get	look	become	look
change	**be**	lose	make	change	get	take
develop	do	run	happen	talk	look	do
make	**live**	cause	take	try	make	work
get	**think**	do	**have**	grow	increase	try
look	**watch**	**live**	look	begin	take	get
take	**stand**	turn	talk	experience	fight	come
fight	wait	come	fight	work	go	say
turn	deal	force	play	take	move	give
work	fight	get	say	fall	occur	follow
strive	listen	grow	work	**hope**	prepare	leave
go	work	happen	**live**	make	**stand**	play
deal	become	**have**	**suffer**	move	change	talk
sit	**lie**	use	grow	plan	deny	wait
study	play	work	begin	receive	develop	provide

Table 7.1. The top 20 verbs in the progressive across all 6 corpora and the BNC_written

A closer look at the progressive uses of the verb *be* in the four native corpora reveals that in the overwhelming majority of the cases it functions as an auxiliary verb in passive constructions, as illustrated in examples 7.1 and 7.2 taken from *LOCNESS*_br and *LOCNESS*_us respectively:

7.1. Motorways and other transport links **are** constantly **being extended**, widened and slowly turning the country into a concrete jungle yet it is only trying to cope with the increase in traffic, we are our own enemy! <Transport 01>

7.2. They also feel as if their First Amendment right **is being infringed upon**. <*ICLE*-US-MRQ-0034.1>

Example 7.1 can thus be classified as a serial state after Huddleston and Pullum's (2005: 166) description of the use of the progressive with stative verbs to express a kind of duration which "tends to be accompanied by an emotive overtone, usually of disapproval, when emphasised by [...] adjuncts as *always, continually, constantly, everlastingly, forever, perpetually*" (Huddleston and Pullum 2005: 166). Example 7.2 expresses a temporary state, which presupposes that the First Amendment is normally not infringed upon and the described situation is only temporary and can easily change. A note of caution is in order here: the inherent lexical aspect of many of the progressive passives is difficult to determine, since the main verbs in the passives are often accomplishment and achievement verbs rather than stative verbs (e.g. *extend, infringe*) and convey "the result state of a situation" (Eriksson 2008: 185). Eriksson (2008: 185) notes that such passives can be analysed in terms of the resulting states of the main verbs after Biber et al.'s (1999: 936) description of the meanings of the passive and points out that learner uses of the progressive passive are thus difficult to classify in terms of their targetlikeness (cf. Eriksson 2008: 185)[56].

In contrast to the variable and not always straightforward interpretation of the progressive passive, active uses of the progressive with stative verbs in the native corpora mostly fall under Huddleston and Pullum's (2005: 167) categories of admissible combinations of the progressive with stative verbs such as temporary states with or without a negative emotional overtone (e.g. *She is cycling to work this week*), waxing and waning situations (e.g. *He is looking more like his father every day*) and agentive activity (e.g. *He is being tactful*) (cf. Huddleston and Pullum 2005: 167). Examples 7.3 and 7.4, taken from *LOCNESS*_us and *FLOB*_F respectively illustrate Huddleston and Pullum's (2005: 167) acceptable uses of the progressive with stative verbs:

7.3. What these parents are failing to understand is that youngsters **are having** sexual relationships at a much younger age, compared to when they were youngsters. <*ICLE*-US-MRQ-0028.1>

7.4. The President **is hoping** to exploit these splits further in the March regional elections and the parliamentary elections in 1993. < F18 111-113>

---

[56] A detailed discussion of learner misuse of the progressive with stative and other verb types will be provided in chapter 8

Here, example 7.3 conveys both agentive activity and a possible disproval on part of the writer with regard to sexual relationships at an early age, whereas example (7.4) refers to a temporary, ongoing state of mind of the president that could possibly change. However, in contrast to the native uses of the progressive with active stative verbs, learner uses of the progressive with active stative verbs tend to deviate from the native uses given above. To illustrate, both Bulgarian and German EFL learners use the progressive to refer to generic or habitual states that are often neither temporary nor waxing and waning, as shown in examples 7.5 and 7.6 taken from *BUCLE* and *GICLE*:

7.5. Unfortunately I doubt that anyone who **was living** twelve years ago had the some humble dreams. *<ICLE*-BG-SUN-0205.1>

7.6. Totally damaged cars, or what **is being left** of them, heaps of shattered glass, puddles of blood on the road and five badly injured, dead bodies lying in or in front of the involved cars. *<ICLE*-GE-AUG-0027.1>

In example 7.5 the Bulgarian learner refers to a static past situation which does not imply any temporariness, dynamicity or an ongoing progress; in example (7.6) the German learner refers to a single occurrence which is the resultative product of a car crash and presupposes no change of state or temporariness. Thus, both learner corpora feature a number of examples of non-targetlike extensions of the progressive to stative verbs which clearly contradict the fourth claim of the Aspect Hypothesis (cf. Andersen and Shirai 1995: 531-532) and which cannot be directly classified in terms of Huddleston and Pullum's (2005) framework of admissible progressives expressing temporary states, agentive activities or waxing and waning situations. These deviating learner examples will be dealt with in greater detail in chapter 8. The following section will present an across-category analysis of the perfect aspect with telic and atelic verbs and will illustrate the 20 most frequent verbs in the perfect across the six corpora under scrutiny.

7.2. The Perfect and the Influence of Inherent Lexical Aspect

The present section deals with a contrastive overview of the distribution of lexical verb types marked for the perfect aspect in learner and native writing. Similar to the across-category analysis of the progressive verb phrases, all perfect verb phrases extracted from the six corpora *BUCLE*, *GICLE*, *LOCNESS*_br, *LOCNESS*_us, *FLOB*_F and *FROWN*_F were

carefully examined together with their larger context in order to determine their inherent lexical aspect. Since the Aspect Hypothesis does not differentiate between achievement and accomplishment verbs with regard to past and perfective grammatical markings (cf. Andersen and Shirai 1995: 531), the present across-category analysis will group accomplishments and achievements together into 'telic perfect verb phrases' as opposed to 'atelic perfect verb phrases' which will include states and activities. Figure 7.3 illustrates the twofold categorisation of the perfect aspect into telic and atelic perfect verb phrases.
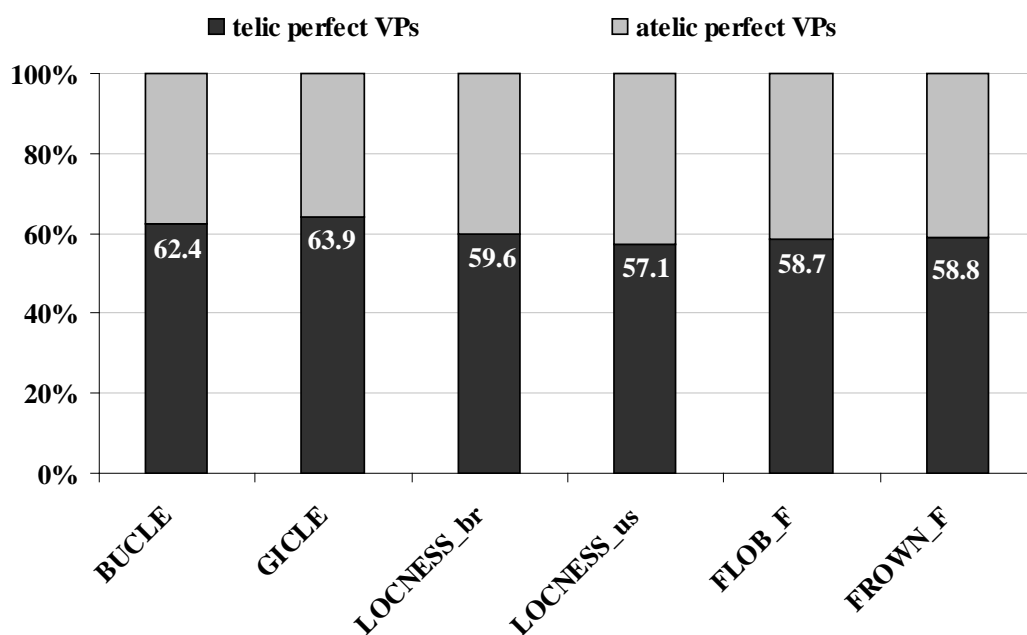
Figure 7.3. Distribution of the perfect verb phrases across lexical verb types

The overall distribution of telic and atelic verb phrases marked for the perfect aspect reveals a less interesting picture than the distribution of telic and atelic progressive verb phrases: on average, the novice and expert native corpora feature less telic verbs in the perfect than the two learner corpora, *GICLE* having the highest ratio of telic verbs in the perfect and *LOCNESS*_us the lowest. This finding suggests a possible preference of both learner groups for accomplishment and achievement verbs in the perfect, much in accordance with the first claim of the Aspect Hypothesis (cf. Andersen and Shirai 1995: 531); however, the statistical significance test shows no significant differences between the learner and native corpora in this respect except for the comparison between *BUCLE* and *GICLE* with *LOCNESS*_us, where both learner corpora feature a slight overuse of telic verbs in the perfect (p<0.05).

Likewise, a closer look at the 20 most frequent verbs in the perfect in the six corpora in comparison to the 20 most frequent perfect verbs in the written section of the BNC reveals no particular trend with respect to learners' preference for telic verbs in the perfect: both *BUCLE* and *GICLE* feature the same or even a lower number of telic verbs among the top twenty perfect verbs (highlighted in the darker cells in table 7.2) than the native corpora and the written part of the BNC.

BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB_F*	*FROWN_F*	BNC_written
be	be	be	be	be	do	be
become	become	make	become	become	be	make
make	have	become	change	see	try	give
change	change	lead	have	have	come	see
turn	see	have	make	take	say	take
learn	do	come	see	make	become	use
do	show	cause	come	come	get	do
lose	come	create	hear	go	look	know
have	find	give	do	give	make	find
try	happen	see	take	show	increase	call
achieve	make	bring	begin	survive	take	get
come	reach	increase	use	do	fight	base
choose	forget	mean	create	bring	go	say
take	give	show	develop	lead	move	show
dream	lose	begin	find	lose	occur	go
give	hear	benefit	give	spread	prepare	hold
create	take	change	lead	fall	stand	set
develop	learn	do	leave	increase	change	come
see	get	find	prove	turn	deny	leave
study	go	take	show	announce	develop	tell

Table 7.2. Top 20 verbs in the perfect in the 6 corpora and the BNC

Therefore, the across-category analysis of lexical verb types across the two aspect forms reveals that the distribution of lexical verb types across perfect forms in the learner and native corpora does not support the claims of the Aspect Hypothesis to the same extent as the distribution of lexical verb types across progressive verb phrases – although both Bulgarian and German EFL learners deviate from *LOCNESS*_us in terms of their preference for telic verbs in the perfect, this learner deviation is by far not as significant as the learner deviation from the novice and expert native corpora as in the case of the progressive. Nevertheless, a slight tendency for both learner groups to use more telic verbs in the perfect than the native novice and expert writers can still be identified; on the other hand, this may be due to learners' more limited vocabulary and higher use of several fairly basic and highly frequent lexical verbs, which at the same time happen to be mostly telic, e.g. *become*, *give*, *make* and *take*. In addition, an across-category analysis of the relationship between the four types of

inherent lexical aspect and one particular grammatical morpheme is always sensitive to the overall distribution of telic and atelic verbs in a corpus, which may not always be balanced – Bardovi-Harlig states that telic verbs are usually more common than atelic verbs (cf. Bardovi-Harlig 2000; 2002), which is also confirmed to a certain extent by the fact that the overwhelming majority of lexical verbs in the perfect in all six corpora in this analysis, as well as in the written part of the BNC, are all telic.

In view of the possible differences between the vocabulary range in the learner and native corpora suggested above, the overall distribution of lexical verbs in the progressive and the perfect in the learner and native corpora have been compared in terms of their type-token ratios (TTR)[57]. The comparison between the six corpora measured by means of a standardised type-token ratio in percent is shown in figure 7.4.
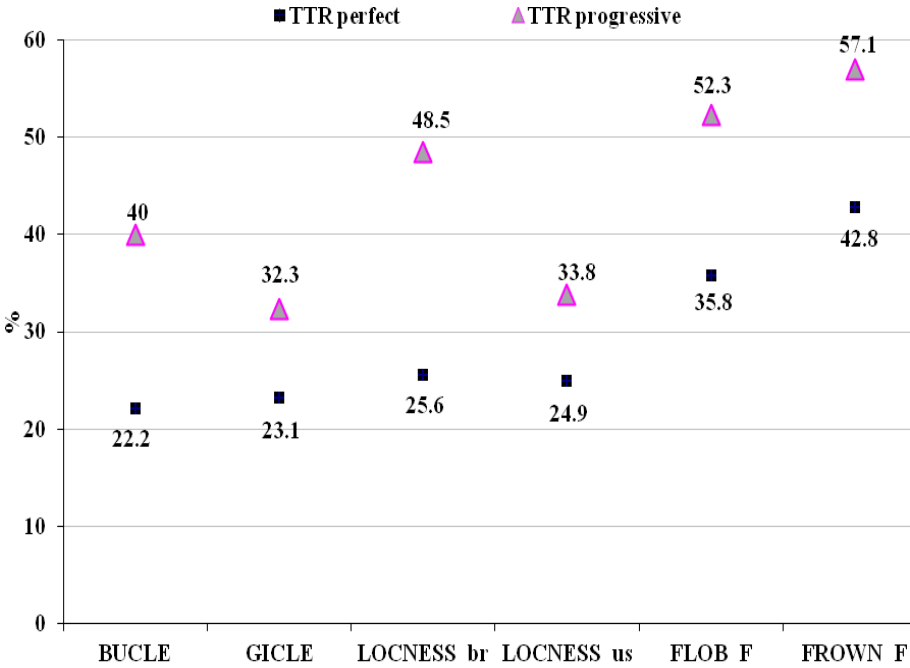
Figure 7.4. Type-token ratio of the progressive and the perfect VPs in all six corpora

The comparison between the TTRs in the six corpora presented above reveals interesting insights with regard to the overall distribution of lexical verbs in the progressive and the perfect: whereas the expert native corpora *FLOB*_F and *FROWN*_F feature the

---

[57] The type-token ratios for the progressive and the perfect were calculated using the refined token counts for progressive and perfect verb phrases in the learner and native corpora (on the basis of the initial Wmatrix count, refined by a subsequent Wordsmith count, filtered and exported to Excel) and a manual extraction of the lexical verb types

highest TTR ratios for both the progressive and the perfect, the learner corpora feature the lowest TTR ratios, with the exception of the TTR for the progressive in *BUCLE*. A low TTR corresponds to a limited number of verb types in the progressive and the perfect that are repeated many times i.e. amount to many tokens; thus, a low TTR can be assumed to be as a general sign of a less diverse vocabulary and possibly a corresponding lower proficiency of the learners (cf. Axelsson and Hahn 2001: 23)[58]. However, in addition to Bulgarian and German EFL learners' proficiency as a possible factor influencing the TTR in the perfect and the progressive, there are several further important factors which need to be mentioned here: native speakers' writing competence, the overall number of progressive and perfect tokens in the corpora, as well as the topic variation in the learner and native corpora. The first factor can account for the differences between the novice and expert native corpora: both *FLOB*_F and *FROWN*_F feature higher TTR ratios for the progressive and the perfect than *LOCNESS*_br and *LOCNESS*_us, which may in turn reflect the fact that *FLOB*_F and *FROWN*_F comprise published written material produced by expert writers, whereas *LOCNESS*_br and *LOCNESS*_us consist of student essays written by high school and university students who are far less competent in writing than journalists and writing experts. On the other hand, the TTRs for the progressive in all six corpora are higher than the TTRs for the perfect, which may be explained with the fact that there are far more perfect verb phrases or perfect verb tokens than progressive verb tokens on average (in an average ratio of 2:5), so that the relatively high number of perfect tokens in contrast to the relatively low number of progressive tokens directly results in a low TTR ratio for the perfect and a high TTR ratio for the progressive. This tendency is further confirmed by the fact that both *BUCLE* and *LOCNESS*_br have unusually high TTR ratios for the progressive; at the same time, these two corpora feature comparatively few progressive tokens altogether (see the previous chapter). Still, there are more progressive tokens in *BUCLE* than in *LOCNESS*_br, *FLOB*_F and *FROWN*_F; the TTR ratio in *BUCLE* is nevertheless lower than the TTR ratios in these three native corpora. Similarly, there are more progressive tokens in *GICLE* than in *BUCLE* and in all four native corpora; nevertheless, the TTR ratio for the progressive in *GICLE* is still lower than the TTR in these corpora.

---

[58] Axelsson and Hahn (2001: 14) identified a TTR ratio of 34.5 for the progressive on the basis of a 79,562-word sample of the German component of *ICLE*; the slight deviation from the TTR obtained in the present study can be explained with the bigger corpus, as well as with the manual refinement of the data in terms of the filtering of 'proper' progressives of all –*ing* forms.

In sum, both Bulgarian and German EFL learners seem to use a fairly limited number of high-frequency verbs with both aspect forms, deviating from the corpus-based native norm established by the TTR ratios of the novice and expert native corpora. At the same time, a closer look at the novice and expert native corpora suggests that proficiency is not the only factor at play here – both novice corpora *LOCNESS*_br and *LOCNESS*_us have lower TTRs for both aspect forms than the expert native corpora *FLOB*_F and *FROWN*_F. Considering the fact that the expert native corpora *FLOB*_F and *FROWN*_F are smaller than *LOCNESS*_us and feature less progressive and perfect tokens altogether, a comparison between the TTRs in *LOCNESS*_br and *FLOB*_F/*FROWN*_F (largely comparable in both size and number of progressive and perfect tokens) should be able to reveal true differences in the vocabulary range of novice and expert native writers. Indeed, this comparison shows a steady cline in the TTRs for the progressive and the perfect from the novice native to the expert native corpora. A possible explanation for this cline may be offered by the fact that similar to EFL learners, novice native writers also use a limited number of highly frequent lexical verbs in the progressive and the perfect due to their insecurity and lack of experience in expository writing. Yet, a further reason behind this novice-expert native cline may be due to differences in the corpus design and topic variation in the novice and expert native corpora – whereas *LOCNESS*_br is based on only 10 different essay prompts, *FLOB*_F features 44 different text samples written on 44 different topics and *FROWN*_F even 48 (see chapter 5). The differences between the number of topics and text samples in the novice and expert native corpora also support the *LOCNESS*_br < *FLOB*_F < *FROWN*_F cline, where the native corpus with the smallest number of essay topics (*LOCNESS*_br) has the lowest TTR, whereas the native corpus with the highest number of text samples (*FROWN*_F) has the highest TTR for both the perfect and the progressive. To conclude, the TTR comparison between the learner and native corpora not only helps to account for the possible differences in vocabulary range and learner proficiency, but it also brings to light individual factors such as writing competence and corpus-related factors such as corpus design and topic variation. Having considered the distribution of lexical verb types across aspect forms in learner and native writing with regard to the influence of inherent lexical aspect as postulated by the Aspect Hypothesis, the next section will focus on the distribution of aspect forms across main and subordinate clauses in order to examine the claims of the Discourse Hypothesis (see chapter 3).

7.3. Distribution of the Progressive and the Perfect across Clause Types

The present section discusses the distribution of progressive and perfect verb phrases in learner and native writing with respect to the central claim of the Discourse Hypothesis that perfective aspect is found primarily in foregrounded clauses which convey "dynamic, kinetic events" (Hopper 1989: 216), whereas imperfective aspect (i.e. the progressive) is primarily found in backgrounded clauses which comment, explain or enhance the main narrative line (cf. Hopper 1989: 213 – 214). Since "verbs will as a rule be interpreted as backgrounding their events if they occur in *subordinate* temporal clauses" (Couper-Kuhlen 1994: 231, original emphasis), the distribution of progressive and perfect verb phrases has been analysed in terms of their occurrence in main and subordinate clauses first, in order to establish possible interdependence between the choice of aspect forms and foregrounded and backgrounded clauses in the learner and native corpora. So far, only a few studies have focused on the distribution of aspect markers across foregrounded and backgrounded clauses, calling for further research on the distribution of morphological markers in backgrounded clauses in particular, which have been identified as "promising for the study of perfect and progressive" (Bardovi-Harlig 1995: 285). Figure 7.5 demonstrates the distribution of progressive verb phrases across main and subordinate clauses in the two learner and four native corpora.

Figure 7.5. Distribution of progressive verb phrases across main and subordinate clauses

The distribution of progressive verbs forms across main and subordinate clauses reveals a consistent trend with respect to both learners' and native speakers' preference to use progressive verb phrases in main clauses: all six corpora feature more progressives in main than in subordinate clauses (over 50% of all progressives are found in main clauses, 55.4% on average), the expert American corpus *FROWN*_F having the highest ratio of progressive verb phrases in main clauses (close to Mindt's (2000: 265) figure of 60% on average), and the novice American corpus *LOCNESS*_us the lowest. The statistical significance test shows no significant differences between the learner and native corpora, thus supporting previous observations on learner performance stating that the strength of the correlation between tense-aspect morphemes and discourse grounding diminishes proportional to the rising proficiency of the learners (cf. Bailey 1989; Bardovi-Harlig 1995), and at the same time confirming Biber et al.'s and Mindt's claim that the progressive occurs predominantly in main clauses (cf. Biber et al. 1999: 461; Mindt 2000: 265). Nevertheless, a slight trend for both learner groups and for the novice native writers of *LOCNESS*_br and *LOCNESS*_us to use more progressives in subordinate clauses than the expert native writers of *FLOB*_F and *FROWN*_F can still be established, suggesting a possible discourse influence on the use of the progressive to express backgrounding situations necessary for the understanding of the main story line in accordance with the claims of the Discourse Hypothesis. Such uses are illustrated in examples 7.5 to 7.8, taken from *BUCLE*, *GICLE*, *LOCNESS*_br and *LOCNESS*_us respectively:

7.5. Fifteen years ago, when Bulgaria was still under the communist regime, people were constantly reminded that they **were building** an equal society, in every respect. <*ICLE*-BG-SUN-0182.1>

7.6. A thief or a burglar is certainly not interested in murdering anybody but if he can be quite sure that the person who might detect him **is carrying** a weapon he would consider himself a fool if he did not do so himself. <*ICLE*-GE-SAL-0001.4>

7.7. Although Margaret Thatcher **is** no longer **leading** the country, the position of most British politicians on the European issue appears extremely ambiguous. <*ICLE*-BR-SUR-0005.3>

7.8. She shows how women **are being degraded** through the publications of the Sports Illustrated swimsuit edition. <*ICLE*-US-MRQ-0020.1>

In the above examples, the past and the present progressive are used in relative and adverbial subordinate clauses which either comment on, enhance or explain the major story line rendered in the main clauses: thus, the fact that Bulgarian people (example 7.5) are

supposed to be building an equal society is a point which enhances the information given in the main clause that they are constantly reminded of that fact; the fact that a thief (example 7.6) might be detected by a person carrying a weapon is a piece of information which is necessary to explain why the thief's use of his or her weapon would be justified in such a case. Similarly, example 7.7 explains first that Margaret Thatcher was sceptical with regard to Britain's involvement in the EU, before proceeding with the main narrative line on Britain's ambiguous position towards the EU; example 7.8 is an embedded relative clause which comments on the representation of women in men's magazines as a juxtaposition to the main narrative on women's emancipation. Even if learners and novice native writers use slightly more progressive verb phrases in subordinate and backgrounded clauses which refer to a point that has already been introduced as part of the given information in the foreground (cf. Dry 1983: 32-33), the differences are not substantial enough to lend full support to the Discourse Hypothesis.

Somewhat more significant are the differences between the learner and native corpora in the case of the distribution of perfect verb phrases across main and subordinate clauses, illustrated in figure 7.6.
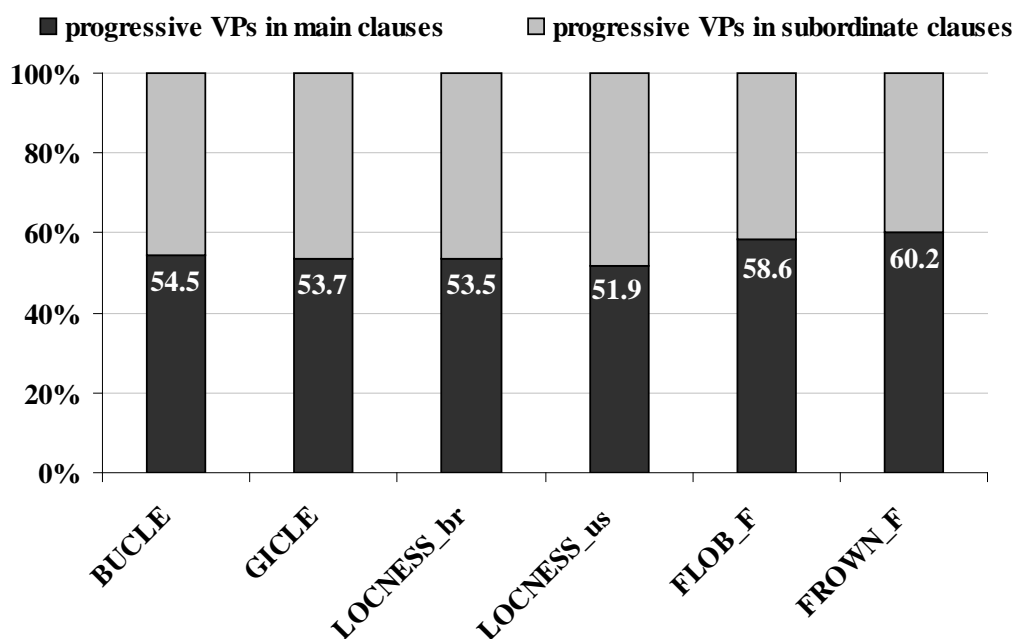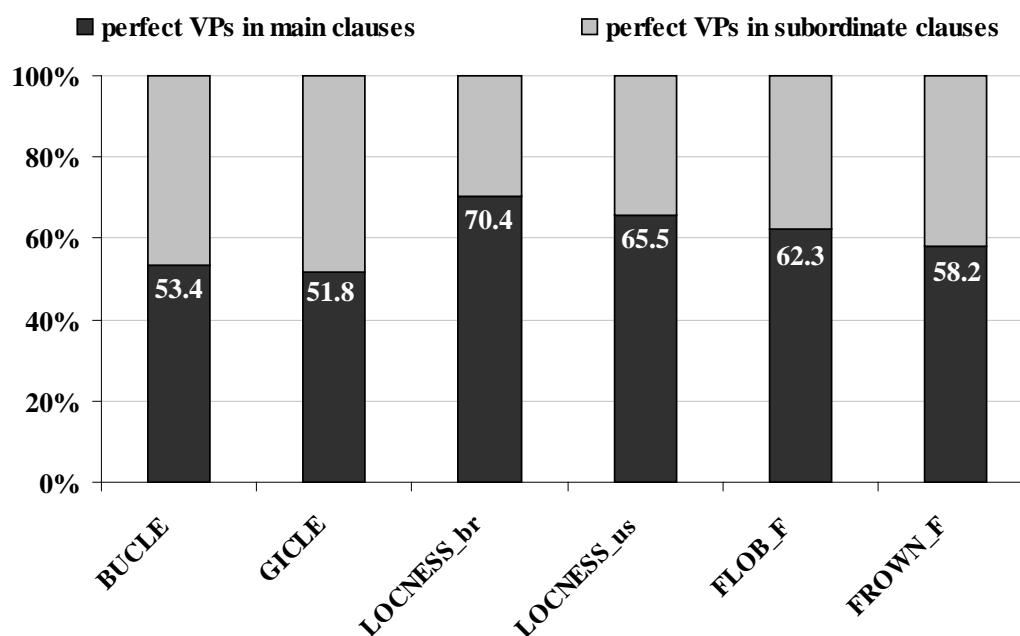

Figure 7.6. Distribution of perfect verb phrases across main and subordinate clauses

The distribution of perfect verbs forms across main and subordinate clauses reveals a different trend with respect to learners' use of perfect verb phrases in main clauses: in contrast

to the relatively uniform distribution of progressive verb phrases in main clauses in all six corpora, the distribution of perfect verb phrases is much less uniform: thus, both *BUCLE* and *GICLE* feature considerably lower ratios of perfect verb phrases in main clauses in comparison to the other four native corpora, *GICLE* having the lowest ratio (51.8%) and *LOCNESS*_br the highest (70.4%). The statistical significance test shows significant differences between both learner corpora and all three native corpora *LOCNESS*_br, *LOCNESS*_us and *FLOB*_F at the level of $p<0.05$ (there are no statistically significant differences between either learner corpus or *FROWN*_F). These findings indicate that although more than half of the perfect verb phrases in Bulgarian and German EFL writing are found in main, rather than in subordinate clauses, thus confirming e.g. Mindt's observation that the perfect occurs predominantly in main clauses (cf. Mindt 2000: 229), there is nevertheless a strong tendency for both Bulgarian and German EFL learners to use perfect verb phrases in subordinate clauses to a much greater extent than novice and expert native writers. Since over 70% of all past perfect forms in *GICLE* were found to be part of subordinate clauses, and since *GICLE* is the corpus with the lowest ratio of perfect verb forms (present and past) found in main clauses, the use of present perfect verb forms in subordinate clauses in *GICLE* merits a closer investigation.

Like the past perfect forms, the majority of all present perfect verb forms in subordinate clauses are found in *wh-*, *that-* and zero-pronoun relative clauses, as illustrated in the following two examples:

> 7.9. If one looks in music shops, one sees thousands of others **who have flooded** the marked with their productions. <*ICLE*-GE-AUG-0006.2>

> 7.10. I think **I have been** born this way. <*ICLE*-GE-AUG-0050.3>

Example (7.9) is part of a *wh*-clause, whereas example (7.10) is introduced by a zero-relative pronoun after 'I think' [59]. In addition to the relative clauses illustrated above, a substantial number of present perfect forms in subordinate clauses are introduced by adverbials and subordinating conjunctions such as *when*, *once* and *as*, as shown in examples 7.11 – 7.12:

> 7.11. Every time when I **have arrived** the middle of the street, suddenly a car seems to come out of nowhere and drives right in my direction. <*ICLE*-GE-AUG-0053.1>

---

[59] Non-targetlike examples like example (7.12), where the simple past would have been the targetlike form will be discussed in greater detail in the next chapter.

7.12. As **I have joined** the army for 12 months I have experience enough to assume that not only me but also about 90% of the boys whom I **have met** there would be willing and would be glad to make use of their "military knowledge" for "humanitary help" even if they run the risk to be shot. <*ICLE*-GE-AUG-0077.1>

In example (7.12), the conjunction *as* is once again used ambiguously – similar to the uses of *as* in combination with past perfect forms discussed in the previous chapter, the use of *as* in combination with the present perfect form shown above can either be interpreted as synonymous to *since* and thus used to introduce a reason, or as a false friend of the German conjunction *als*, which would have specified the temporal circumstance of joining the army in L1 German: the German translation of example (7.12) into *als ich der Armee beitrat* seems plausible here. In addition to the distribution of present and past perfect verb forms in relative and adverbial subordinate clauses, a small number of present perfect verb phrases occur in if-clauses like the following two examples:

7.13. If **you've told** a secret to a so-called friend and he had nothing else to do but run to some other people and tell them about it, than you can surely live better without such friends. <*ICLE*-GE-AUG-0060.1>

7.14. Such machines should automatically solve the problems with the repairing of satellites if **they've had** a breakdown high up in the sky. <*ICLE*-GE-AUG-0047.1>

Here, the present perfect is part of an *if*-clause rendering two possible and open conditions, which can be interpreted either as likely (e.g. the author's friends giving their secrets away or satellites breaking down are two likely conditions) – in which case the present tense in the *if*-clause in combination with the *will*-future in the main clause would have been the traditional option as defined in empirical grammars of English (e.g. Quirk et al. 1985: 1010) and thus the arguably more acceptable option, or alternatively as unlikely and hypothetical conditions, in which case the simple past in the *if*-clause in combination with the past modal form *would* would have been the more acceptable option. On the basis of the propositions in the two sentences illustrated above, it can be safely assumed that the first version (the conditions being open and likely rather than hypothetical) is the more probable version of the two, in which case the present tense and not the present perfect would have been the more acceptable choice[60].

---

[60] A detailed discussion of the non-targetlike uses of the perfect will be offered in the next chapter

Similar to the *GICLE* learners, the *BUCLE* learners also use the perfect (present and past) in subordinate clauses to a much greater extent than the native writers of *LOCNESS_*br, *LOCNESS_*us and *FLOB_*F. The majority of all present and past perfect verb forms in subordinate clauses are found in *wh-*, *that-* and zero-pronoun relative clauses, as well as in adverbial clauses. A number of present perfect forms were also found in *if-*clauses, as demonstrated in the following examples:

7.15. If **you have dreamt** of something and want it very strongly indeed, then the realisation of your dream becomes your aim in life. *<ICLE*-BG-SUN-0044.1>

7.16. Because if **people have not realised** their perishableness, they would not have striven for making a good job of their lives […]. *<ICLE*-BG-SUN-0038.1>

Example 7.15 refers to an open condition which is likely for a daydreamer (i.e. the topic of the essay), whereas example (7.16) refers to a hypothetical condition which is highly unlikely (people have long realised that they are mortal). Therefore, the more acceptable form in the *if-*clause in the former example would have been the present tense, rather than the present perfect followed by the *will-*future in the main clause, whereas the acceptable (if not only possible) form in the second hypothetical example would have been the past perfect followed by the perfect modal form *would have been*.

To summarise, given that the verbs that occur in subordinate temporal clauses are part of the background of the main story line (cf. Couper-Kuhlen 1994: 231), the distribution of the progressive and perfect verb forms across main and subordinate clauses in learner and native writing patterns differently with regard to the claims of the Discourse Hypothesis: whereas progressive forms are distributed almost uniformly across all six corpora, occurring mostly in main clauses (and thus also confirming Couper-Kuhlen's (1994: 229) findings that the progressive occurs predominantly in main clauses in "flagrant violation" with the rules postulated by the Discourse Hypothesis, together with Biber et al.'s (1999: 461) and Mindt's (2000: 256) general findings on the progressive), the distribution of perfect forms across main and subordinate clauses in both Bulgarian and German EFL writing shows greater deviations from the native corpus-based norm. Even though both learner groups use more perfect forms in main than in subordinate clauses, they still show a stronger preference for the perfect aspect in subordinate clauses in comparison to the native speakers in the present sample. A higher proportion of perfect forms in subordinate clauses may be explained with a stronger correlation between the perfect (and the past perfect in particular) and the background of a

story: thus, present and past perfect forms are favoured by Bulgarian and German EFL learners (rather than by native speakers) to render old information, before the new, foregrounded information is delivered via other verb forms.

These findings are interesting since they show that whereas progressive verb forms occur predominantly in main clauses in both learner and native writing (cf. Couper-Kuhlen 1994), the distribution of perfect verb forms across main and subordinate clauses varies according to whether the writers are EFL learners or native speakers of English. In the latter case, the Discourse Hypothesis seems confirmed to a greater extent than in the case of the learners. This observation needs further clarification: it seems that the correlation between the choice of aspect form and grounding is stronger in the case of the perfect aspect than in the case of the progressive aspect, as well as stronger in the case of the three native corpora *LOCNESS_*br, *LOCNESS_*us and *FLOB_*F, than in the learner corpora and in *FROWN_*F. However, a word of caution is in order here: similar to the limitations of the across-category analysis of lexical aspect types in the progressive and the perfect discussed in the previous section, an analysis of the clauses containing only progressive and perfect verb forms can likewise influence the results, since they will be indicative only of these two forms and not of the distribution of all perfective and imperfective markers in the corpora – it may well be the case that a comprehensive analysis of all clauses in the corpora would yield different results with respect to learners' and native speakers' preference for tense-aspect forms in main and subordinate clauses. Unfortunately, such an analysis goes beyond the scope of the present study. The next section will examine a further lexical factor accompanying the use of aspect forms in learner and native written English: the use of temporal adverbials modifying progressive and perfect verb phrases.

7.4. Adverbial Modification of the Progressive and the Perfect

Both the progressive and the perfect aspect commonly occur in the company of a number of adverbs which modify the actions and events they describe – the present section focuses on this co-occurrence, and in particular on the co-occurrence of the two aspect forms with adverbials of time and temporal adverbial phrases in learner and native writing. Figure 7.9 illustrates the ratios of the temporally-modified progressive verb phrases in the six corpora.
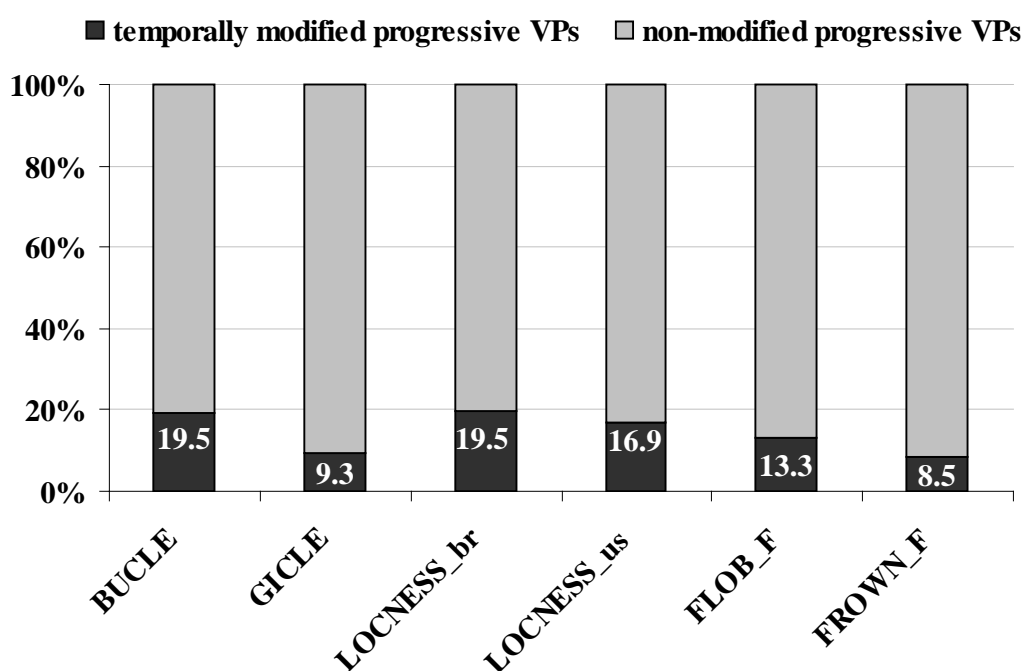
Figure 7.7. Temporal modification of the progressive VPs in %

The ratios for the learner and native corpora show a rather diverse picture: whereas *BUCLE*'s ratio for temporally modified progressive verb phrases is the highest one, together with *LOCNESS*_br's ratio, deviating significantly (p<0.001) from both *GICLE*'s and *FROWN*_F's ratios, there are no significant differences between *BUCLE* and the other three native corpora. In contrast, *GICLE*'s ratio is significantly lower than the ratios in the two native novice corpora *LOCNESS*_br and *LOCNESS*_us (p<0.001). This finding suggests that Bulgarian EFL learners possibly emphasise their progressives using temporal adverbials more often than German EFL learners do, which in turn could be due to the fact that *BUCLE* learners use the progressive aspect in a more limited number of ways, most of which strictly temporal, rather than e.g. emotional, matter-of-course, highlighting etc. non-temporal

meanings (cf. Mindt 2000: 256). In comparison, *GICLE* learners, who use the progressive considerably more often than *BUCLE* learners (see chapter 6), use significantly less temporal adverbials to modify their progressive verb phrases – this may possibly be due to a more varied use of the progressive by the *GICLE* learners, which also indicates higher proficiency of the German EFL learners. There are no significant differences between the native corpora, apart from *FROWN*_F, which deviates from the other three native corpora in terms of its rather low ratio of temporal modification of the progressive. A closer look at the most frequent adverbs and adverbial phrases occurring with the progressive in the learner and native corpora presents no surprises: in line with Mindt's results (cf. Mindt 2000: 265), the most frequent adverbs accompanying the progressive are *now/nowadays, still, always, already, constantly* and *just*, as illustrated in table 7.3:

BUCLE	*GICLE*	*LOCNESS*_br	*LOCNESS*_us	*FLOB*_F	*FROWN*_F
constantly	always	constantly	constantly	now	now
still	constantly	already	now	already	still
now	still	still	still	more and more	already
always	already	now(adays)	already	just	always

Table 7.3. The 4 most frequent adverbs occurring with the progressive in the six corpora

Rather more interesting are the findings with respect to the temporal modification of the perfect and the present perfect in particular. Since temporal adverbials like *since*, *for*, *just* etc. have traditionally been employed as "trigger words" for the present perfect in EFL teaching contexts (cf. Schlüter 2000; 2002; 2006), and since adverbial modification has been found to be a lot more infrequent than commonly suggested (Schlüter (2002: 313) comments that only 33% of all present perfect verb phrases in different registers of British and American English are temporally modified), a comparison between the rates of adverbial modification in learner and native writing would help to reveal possible teaching-induced effects on Bulgarian and German EFL learners. Figure 7.8 illustrates the rates of temporal modification in the learner and native corpora.
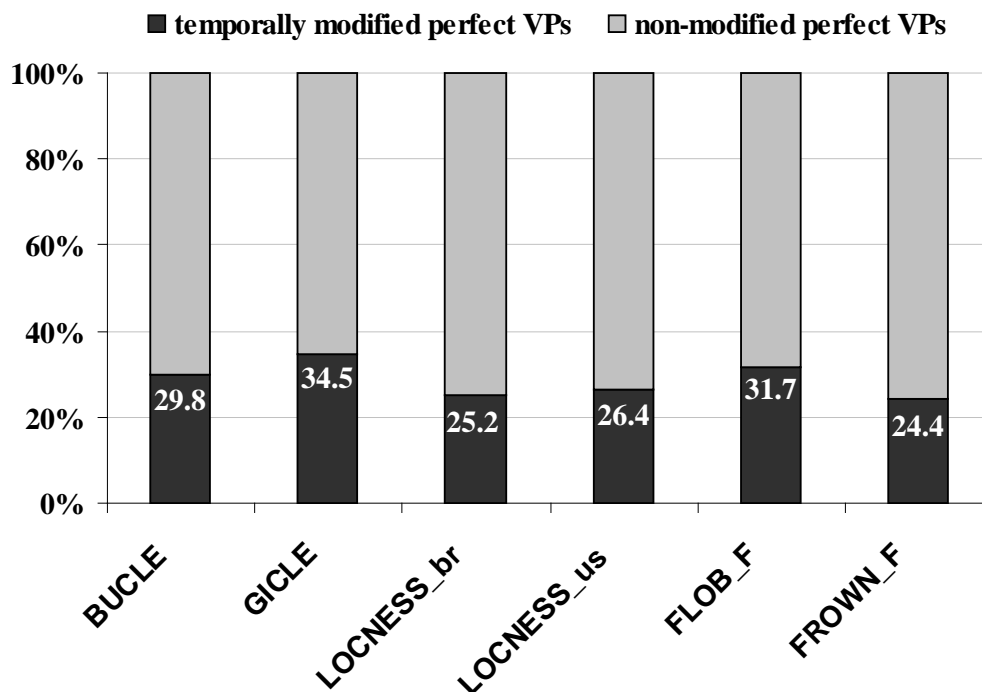
Figure 7.8. Distribution of the temporally modified perfect verb phrases

The average rate of temporal modification of all six corpora is 28.6%, which, on average, is lower than the 33% ratio suggested by Mindt (2000: 313); however, both learner corpora feature higher rates of temporal modification than the novice and expert native corpora (apart from *FLOB*_F), *GICLE* having the highest rate and the expert native corpus *FROWN*_F the lowest. The statistical significance test shows significant differences between *GICLE* and the three native corpora *LOCNESS*_br, *LOCNESS*_us and *FROWN*_F ($p<0.001$); somewhat surprising is the lack of significant differences ($p>0.05$) between *BUCLE* and the native corpora. Nevertheless, this comparison suggests that although both learner groups use more temporal adverbials with the perfect on average, German EFL learners overuse temporal adverbials with the perfect to a greater extent than Bulgarian EFL learners, which indicates a possible teaching-induced bias of German EFL learners to emphasise their perfect verb phrases with a temporal adverbial, thus favouring temporally modified perfect verb phrases to unspecified perfect verb phrases (cf. also Davydova's results (2011: 287). The most frequent adverbs and adverbial phrases co-occurring with the perfect are presented in figure 7.9.
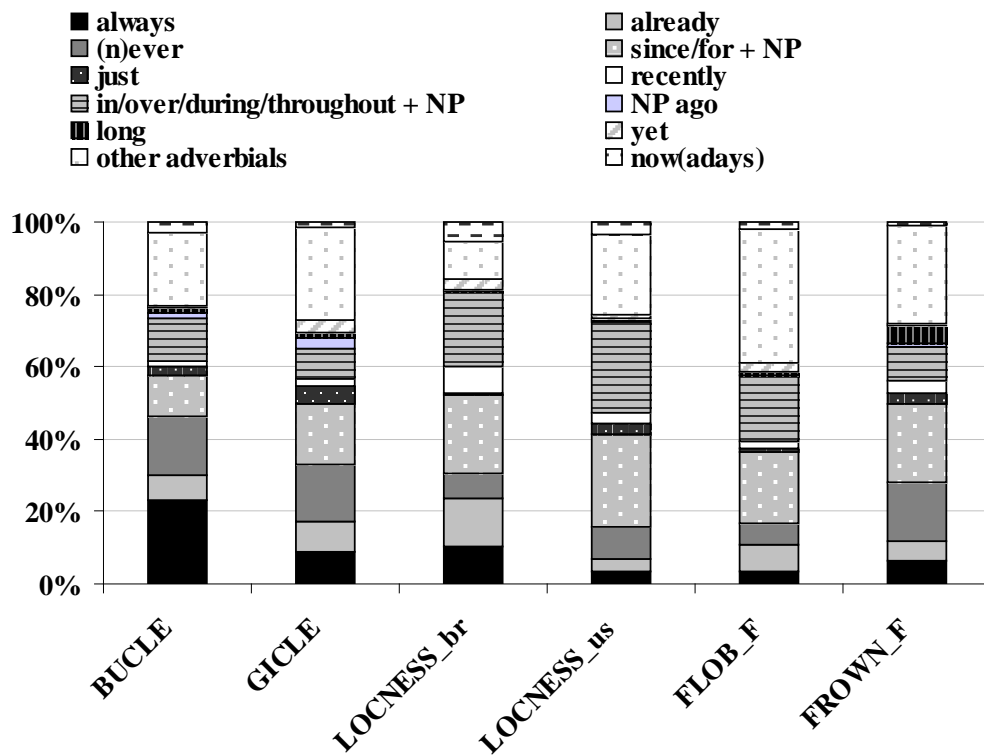
Figure 7.9. Most frequent temporal adverbial phrases

In accordance with Mindt's (2000: 229; 247) and Biber et al.'s (1999: 468) results for the most frequent temporal adverbials with the perfect, the most frequent adverbials in the learner and native corpora (making up for more than 65% of all adverbials) are *already, always, ever, never, since, for*, *just* and *now* – to illustrate, *always* is the most frequent adverbial in *BUCLE*, whereas *since* and *for* in combination with a noun phrase are most common in *GICLE* (cf. also Schlüter's (2002: 313) results with respect to the *(ever)since + temporal noun phrase* construction as the single most common adverbial phrase modifying the present perfect). Somewhat contrary to Biber et al.'s findings that the perfect does not co-occur with temporal adverbial phrases which signal "a clear ending point before the present time", such as *in, during* and *throughout* followed by a noun phrase (Biber et al. 1999: 467), perfect verb forms accompanied by such adverbial phrases (*in, over, during* and *through(out)* + a noun phrase) are fairly common in both the learner and the native corpora, as illustrated in the following examples taken from the native corpora *LOCNESS_*br and *FROWN_*F:

7.16. **In the last 30 years** we have seen the construction of a vast motorway network throughout the U.K. <Transport 03>

7.17. Yet the United Kingdom and the United States **during the last 25 years** have blatantly pursued policies directed at keeping refugees out. <F 09 51-52>

However, a number of instances of the temporal adverbial *ago* preceded by a noun phrase accompanying the present perfect were testified in the learner corpora, but not in the native corpora: to demonstrate, 5 examples (c. 1.5% of all temporal adverbial phrases modifying the perfect) were found in *BUCLE* and 12 (c. 3% of all temporal adverbial phrases modifying the perfect) in *GICLE*. Two such examples (taken from *BUCLE* and *GICLE* respectively) are illustrated below:

7.18. I've only found the answer a few days **ago**. *<ICLE*-BG-SUN-0027.1>

7.19. Our society is not as stable as it has been a generation **ago**. *<ICLE*-FR-ULG-0018.2>

In both cases, the adverbial *ago* signals a temporal frame that has clearly come to an end – several days before the moment of speaking in example 7.18 or a generation before the moment of speaking in example 7.19; nevertheless, the learners use these adverbial phrases in combination with the present perfect, rather than the simple past, which would have been the preferred and targetlike option in these cases. In contrast, the only examples of perfect verb phrases temporally modified by *ago* in the native corpora are two examples of the past perfect, taken from the American corpora *LOCNESS*_us and *FROWN*_F:

7.20. If you had told someone 100 yrs **ago** that you could cut cooking time in half, and not use fire, they would have thought you crazy, or bewitched, or under the influence of heavy drugs. *<ICLE*-US-MICH-0041.1>

7.21. All had been conquered by climbers long **ago**, of course; mountaineers have been coming to this picturesque Alpine village for much more than a century. <F 06 14-16>

The past perfect in the former example depicts a hypothetical situation of a person meeting their dead ancestors, whereas in the latter example it describes the pre-past situation of climbers conquering a village which happened before a series of other events in the past. The use of *ago* in both past perfect examples is not typical and has not been identified as such by the usage-based grammars of English (e.g. Quirk et al. 1985; Biber et al. 1999; Mindt 2000); however, it is still striking that the only two examples come from American English and not British English.

In sum, both learners and native novice and expert writers use a common set of temporal adverbials to modify their present and past perfect verb phrases; however, learners (and in particular German EFL learners) tend to 'overmodify' their perfect forms rather than leaving them unspecified, as well as to combine them with inadmissible temporal adverbials signalling a definite past moment that has come to an end. A brief comparison between the type-token ratios of the temporal adverbials used with the perfect in the learner and native corpora (Figure 7.10) shows a clear rising trend with respect to the lexical range of adverbial phrases: whereas the learner corpora feature the lowest TTRs and thus the least varied adverbial vocabulary, the expert native corpora feature the highest and most varied vocabulary, leaving the novice native writers behind as a middle ground between EFL and expert writing. Thus, similar to the findings on the TTRs of the progressive and the perfect in the six corpora and the corresponding vocabulary range of the learners and native speakers illustrated in sections 7.1. and 7.2, the TTRs of the adverbials co-occurring with the perfect seem again to correlate with writing experience – the more experienced the writers are, the richer their vocabulary.
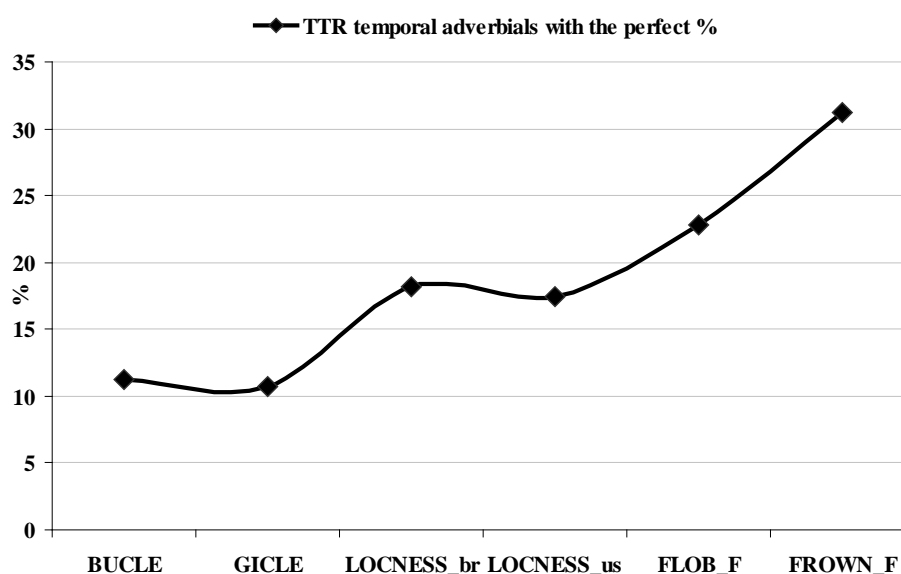
Figure 7.10. Type-token ratios of the temporal adverbials modifying the perfect

The next section will address one final point concerning learners' and native writers' writing experience with regard to their use of the progressive and the perfect in expository writing: it will focus on the tendency of advanced EFL learners to adopt spoken-like features in their academic writing (cf. Gilquin and Paquot 2007), and in particular EFL learners' tendency to use contracted auxiliaries in combination with these two aspect forms.

7.5. Distribution of Contracted Auxiliary Forms with the Progressive and the Perfect

A number of learner corpus studies (e.g. Rayson and Granger (1998); Granger (1998); Lorenz (1999); Aijmer (2002); Gilquin and Paquot (2007) etc.) have identified an "overly oral tone" (Gilquin and Paquot 2007: 2) in advanced EFL learners' writing, which they attribute to advanced EFL learners' overuse of features in writing, which are typical of spoken, rather than written language. The overuse of spoken-like features makes learner writing conceptually oral in nature, locating it closer to the orality end of the orality-scripturality scale proposed by Koch and Oesterreicher (cf. Koch and Oesterreicher 1985). The present section addresses this aspect of advanced learner language, focussing on the variation in use of contracted auxiliary verb forms of *be* and *have* (*I'm, you're, I've* etc.) with the progressive and the perfect in learner and native writing. Figure 7.11 illustrates the rates of contracted auxiliary forms used with progressive and perfect verb forms in the learner and native corpora.
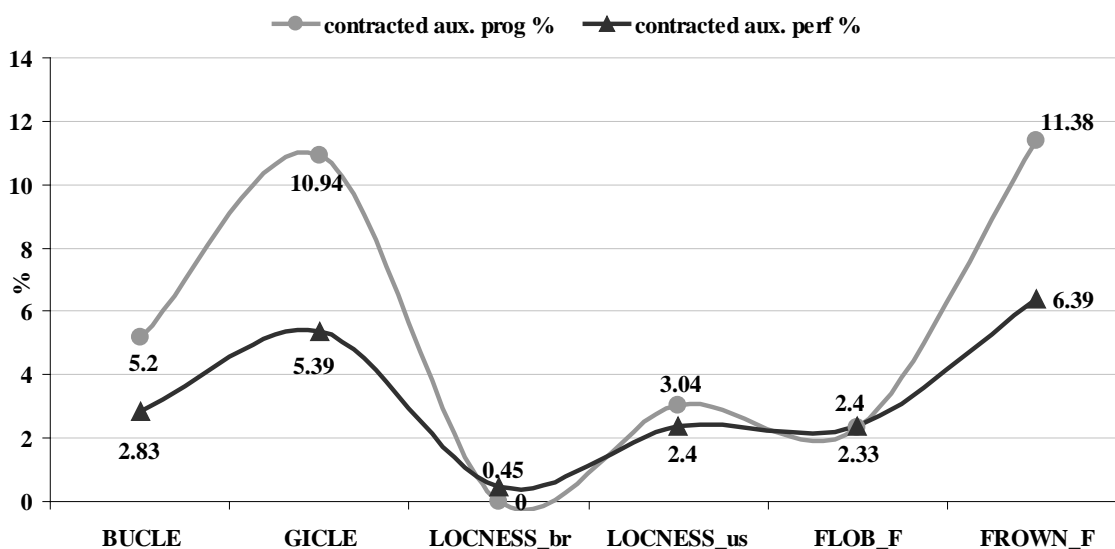
Figure 7.11. Percentage of the contracted auxiliary forms in progressive and perfect VPs

The figure shows that contracted auxiliary forms are distributed fairly unevenly among the learner and native corpora – whereas the American expert native corpus *FROWN*_F features the highest ratio of auxiliary contractions with the progressive and the perfect, the British novice native corpus *LOCNESS*_br has the lowest rates; the learner corpus rates lie in between them, with *GICLE* being closer to *FROWN*_F and *BUCLE* to *LOCNESS*_br. However, the statistical significance test shows a highly significant overuse of contracted auxiliaries with the progressive and the perfect in *BUCLE* in comparison with *LOCNESS*_br

(p<0.001) and, simultaneously, a significant underuse in *BUCLE* in comparison with *FROWN*_F (p<0.05). In contrast¸ there are no differences in the rates of contracted auxiliary forms between *GICLE* and *FROWN*_F; nevertheless, *GICLE* features a significant overuse of contracted auxiliary forms in comparison with the other three native corpora *LOCNESS*_br, *LOCNESS*_us and *FLOB*_F (p<0.001).

One striking regularity in the distribution of contracted auxiliary forms with the progressive and the perfect concerns the patterning of British and American English with respect to lower vs. higher rates of contracted auxiliaries – both British corpora *LOCNESS*_br and *FLOB*_F feature lower contraction rates than the American corpora *LOCNESS*_us and *FROWN*_F; interestingly enough, the learner corpora show a clear resemblance to either the British or the American corpus-based norm: *BUCLE*'s rates are lower and closer to the British corpora, whereas *GICLE*'s rates are higher and closer to the American corpora. This regularity may be explained with the "tendency for spoken language habits to infiltrate the written language" (Leech and Smith 2006: 198), which has been claimed to be particularly typical of American English usage – American English being more prone to adopt spoken-like features in writing than British English, "leading the way" towards colloquialisation of written language (Leech and Smith 2006: 199). Indeed, the lack of occurrences of contracted auxiliaries in the British novice native corpus *LOCNESS*_br reveals a stronger affinity of British novice writers to the prescriptive norm, which advocates the use of the full auxiliary form in writing; at the same time, it may also signal a possible insecurity of the British novice writers to use more varied language, since the expert British writers of *FLOB*_F use more auxiliary contractions than the *LOCNESS*_br writers. Likewise, the German EFL learners, who have had more exposure to spoken English and have been rated as more advanced learners than the Bulgarian EFL learners by independent CEFR raters (see chapter 5), tend to incorporate more contracted auxiliaries in their writing: therefore, both proficiency and target-language exposure seem to be at play here, since the more advanced learners use more colloquial features in their writing. Interestingly enough, the preference for contracted auxiliary forms with the progressive and the perfect in the native corpora does not support the orality-literacy continuum hypothesis (see chapter 2), which would presuppose that the expert writers are also the most literate and thus closest to the scripturality end of the orality-scripturality scale: however, even though the expert writers of *FLOB*_F and *FROWN*_F are much more experienced in writing, they still favour auxiliary contractions to a much greater extent than the inexperienced writers of *LOCNESS*_br and *LOCNESS*_us. This observation is

further corroborated by the fact that the youngest and least experienced students of *LOCNESS*_br do not use any auxiliary contractions at all. In sum, the use of contracted auxiliaries with the progressive and the perfect seems to depend on a multitude of factors, ranging from learner proficiency and writing expertise to the type of native variety of English (British English or American English) under scrutiny. The last section in this chapter offers a brief summary of the results on the lexicogrammatical variation in the use of the progressive and the perfect in learner and native writing.

7.6. Summary

The distributional analysis of progressive and perfect forms in learner and native writing described in this chapter revealed both hypothesised and unpredicted lexicogrammatical differences between the learner and native corpora used in the present study. While the findings of the across-category analysis are in line with the third claim of the Aspect Hypothesis, indicating a certain bias of advanced Bulgarian and especially German EFL learners towards using more activity verbs with the progressive than native speakers, they also contradict the fourth claim of the Aspect Hypothesis which states that "[p]rogressive markings are not incorrectly overextended to stative verbs" (Andersen and Shirai 1996: 533), since both *GICLE* and *BUCLE* feature considerably more "incorrectly overextended" progressive markings than the novice and expert native corpora. In terms of the use of the perfect with accomplishment and achievement verbs, the results are less exciting, since the learner deviation from the native-speaker corpus-based norm is much less significant than the deviation in the case of the progressive: both learner groups show a slight tendency of overusing telic verbs with the perfect than the native novice and expert native writers; at the same time, this deviation seems to be due to learners' more limited vocabulary and higher use of a small number of highly frequent telic verbs more than anything else.

In terms of the distribution of progressives across main and subordinate clauses, a slight trend for both learner corpora, as well as for the novice native corpora *LOCNESS*_br and *LOCNESS*_us to use more progressives in subordinate clauses was identified, thus suggesting that learners and inexperienced native writers are equally biased towards using more progressives in subordinate clauses than expert native writers in writing; thus confirming the claims of the Discourse Hypothesis with regard to the progressive conveying backgrounding information in support of the main story line. Likewise, both *BUCLE* and

GICLE (in particular) feature more perfects in subordinate clauses than the native-speaker novice and expert corpora; a finding which merits a more detailed explanation and which will be addressed in greater detail in the discussion part (chapter 9) of the present study.

Less striking are the results concerning the temporal modification of progressive and in particular perfect verb phrases: as suggested by previous studies (e.g. Davydova 2011), both learner corpora feature higher rates of temporal modification than the native-speaker corpora. At the same time, the adverbs modifying perfect verb phrases are much less varied in the learner corpora than in the native corpora, as well as often inadmissible (e.g. *yesterday*), since Bulgarian and German EFL learners (and to a lesser extent the native novice writers of *LOCNESS*_br and *LOCNESS*_us) stick to a limited number of highly frequent "prototypical" adverbs.

Most striking are the results with respect to Bulgarian and German EFL learners' frequencies of use of contracted auxiliaries with the progressive and the perfect, which show a certain resemblance to either the British or the American corpus-based norm. Whereas *BUCLE*'s rates of use of contracted auxiliaries with the progressive and the perfect are much lower and closer to the British novice and expert corpora, *GICLE*'s rates are higher and thus closer to the American novice and expert corpora, and almost as high as the rates of *FROWN*_F, the corpus with the highest rates of auxiliary contractions with the progressive and the perfect. The reasons behind this resemblance, especially in the light of learner-specific variables such as the target language norm and the amount of target language exposure, will be examined in greater detail in chapter 9 of the present study. The next chapter deals with a qualitative analysis of the use and in particular the misuse of progressive and perfect forms in Bulgarian and German EFL writing, thus drifting away from the CIA framework of interlanguage comparison used so far for the quantitative part of the present study, and moving closer towards a traditional SLA study. Nevertheless, the chapter is based on a novel combination of two methods – a problem-oriented Computer-Aided Error Analysis and a target-like use analysis of learner corpus data, which work together with the goal to detect, quantify and evaluate instances of misuse of progressive and perfect verb forms in advanced Bulgarian and German EFL learner writing.

8. Misuse of Aspect in Bulgarian and German EFL Writing

The previous two chapters dealt with a contrastive comparison between the individual frequencies of use of progressive and perfect forms in learner and native writing, their lexical profile in the light of the Aspect Hypothesis, their distributional patterns in terms of the Discourse Hypothesis, as well as their co-occurrence with temporal adverbials and contracted auxiliaries. The present chapter is dedicated to a more detailed qualitative analysis of the use and in particular the misuse of progressive and perfect forms in Bulgarian and German EFL writing. The chapter starts with a brief overview of the new tendencies in Error Analysis (EA) – Computer-Assisted Error Analysis (CEA) and illuminates some persisting problems of error identification and error correction. Using a novel combination of two methods – a problem-oriented CEA and a target-like use analysis, the use of progressive and perfect forms in advanced Bulgarian and German EFL learners' writing is evaluated and analysed within the temporal context of each essay (sections 8.3 and 8.4). This particular approach allows for an in-depth analysis of "what learners get right as well as what they get wrong" (Ellis and Barkhuizen 2005: 70, see also chapter 5). Finally, section 8.5 presents a possible method of quantifying learner errors in a POS-tagged and error-tagged learner corpus for the purposes of interlanguage comparison.

8.1. Approaching Advanced EFL Learners' Misuse: New Tendencies and Old Problems

The Error Analysis methodology of the early 1970s was a significant breakthrough in SLA theories which focused on the explanation of second language processes well beyond the limitations of Behaviourism, steering SLA research into a brand new direction after Corder's (1967) seminal article on the significance of learners' errors and the subsequent coining of the term "Interlanguage" by Selinker (1972) and the following wave of interlanguage analysis as an autonomous language research (cf. Corder 1967, Selinker 1972; 1976; Gass and Selinker 1994; Kellerman 1997, Sharwood Smith 1994, Ellis and Barkhuizen 2006 etc.). Next to a systematic identification, documentation and description of learners' errors, the major asset of 1970s Error Analysis is the explanation of the sources of learners' errors and their classification into two major types: 1) interlingual errors (errors that result from the influence of the mother tongue, i.e. negative transfer effects) and 2) intralingual errors (errors that are

independent of the mother tongue, i.e. developmental errors) (cf. Ellis and Barkhuizen 2006: 64 – 65). Whereas interlingual errors are the result of negative transfer which is exemplified "by establishing divergences from the norms of the target language" (cf. Odlin 1989: 36), intralingual errors stem from the incompleteness of the L2 system of learners with various mother-tongue backgrounds and proficiency levels who experience the same learning difficulties. The following chart summarises the major types of interlingual and intralingual errors:
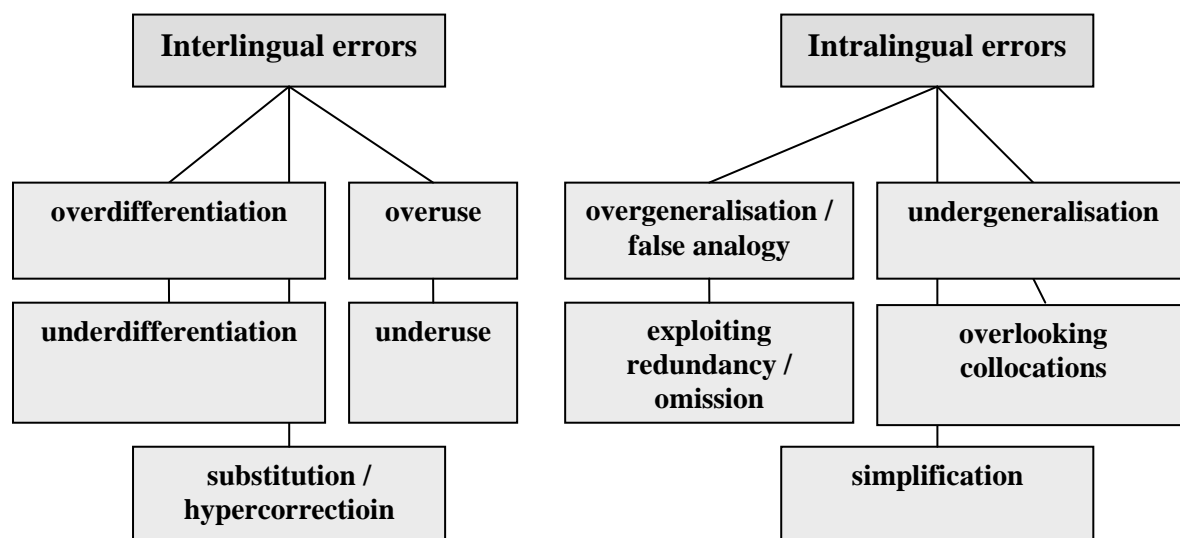
Figure 8.1. Interlingual and intralingual error types (adapted from Ellis and Barkhuizen 2006: 66)

While the boundaries between these error types and in particular their exact sources may well be fuzzy (e.g. a deviation from the norm of the target language may have an interlingual, as well as an intralingual source), researchers argue that "the early stages of language learning are characterized by a predominance of interlingual interference" (Liu 2012: 21), whereas intralingual errors or generalizations are more typical in the later stages of the language learning process (cf. ibid). Moreover, pinpointing advanced learners' interlingual errors with certainty is a major challenge to SLA researchers – Kellerman (1997) tries to explain this difficulty by analysing typologically close languages (e.g. Dutch and English) and their respective learners. He argues that while the small typological distance between the cognates Dutch and English is a definite advantage for learners of either language, often resulting in positive transfer and not in errors, one further important factor is not be neglected – the prototypicality of a feature in the native language. Kellerman maintains that irrespective of the genetic and/or typological closeness between the native and the target language, learners are unwilling to transfer non-prototypical ('non-natural') features from their L1 to the

L2, even if L1 and L2 may have many formal and functional parallelisms (cf. Kellerman 288 – 293). Lastly, he argues that learners with a typologically distant L1s are also less likely to transfer L1 forms into the L2 (cf. Kellerman 1997; Kellerman 1979 in Ellis and Barkhuizen 2006: 65). With regard to the two learner populations in the present study, it can be hypothesised that intralingual (developmental) errors (cf. Dulay and Burt 1974) like under- and overgeneralisation will predominate in advanced Bulgarian and German EFL learners' writing; however, they will predominate **in addition** to the identified ratios of interlingual effects such as e.g. underuse of the progressive and the perfect outlined in chapter 6. Next, the effects of the relative typological distance between Bulgarian and German as L1s and English as L2 are not as straightforward (both German and Bulgarian are typologically distant to English; however, German is much closer, genetically cognate language) and still need to be examined.

Finally, in addition to the interlingual and intralingual errors, a further error type needs to be mentioned here – the so-called errors resulting from the transfer of training (cf. Selinker 1972; 1976; Gass and Selinker 1994). The transfer of training involves an unintended feature in the input "intentionally or unintentionally created by the teacher or textbook" (Sharwood Smith 1994: 37) which is transferred to the L2 with a non-targetlike result. Overemphasis of a particular feature (e.g. the present progressive or the present perfect) in teaching materials and/or in the foreign language classroom may lead to over- or underuse of that feature (cf. Sharwood-Smith 1994: 37): an exemplary study illustrating this phenomenon is Römer's (2005) study on the progressive in German classroom materials – a study which draws attention to the inadequate and often faulty representations of the progressive in textbooks for German EFL learners and their detrimental effect for both teaching and learning. Last, but not least, even though at an advanced level, learners may simply avoid using certain forms they have not fully attained or feel unsure about using (cf. Schachter 1996) – a linguistic behaviour which may not result in overt errors of any of the above mentioned types, but which, however, makes a strong claim about the fossilisation or the partial attainment (cf. Selinker 1972; Odlin 2006) of the second-language system.

Although EA provided SLA researchers in the 1970s with powerful new tools and insights about the nature of learning processes, it suffered from several major drawbacks such as a lack of rigidity, sporadic and anecdotal manner of the collection of data, a strong focus on individual learners, a static picture of L2 learning as well as EA's inability to capture learners'

avoidance of certain linguistic features (cf. Dagneaux et al. 1998: 63). The methodology employed in the present study is Computer-Aided Error Analysis (CEA, see also chapter 5), which is a follow-up method to traditional Error Analysis, but which "has inherited the methods, tools and overall rigour of corpus linguistics" (Dagneaux et al. 1998: 173). CEA classifies the various error types in an error-annotated learner corpus using predetermined error codes that can easily be filtered, extracted and analysed in a structured manner, as well as measured and quantified against the background of corpus data.

Nevertheless, CEA still shares some of the weaknesses of traditional EA like the difficulty of error detection and the multitude of often contradictory target hypotheses behind error correction – while Dagneaux et al. (1998) claim that CEA lacks most of the drawbacks of traditional EA (e.g. the randomness of data, the fuzzy error categories and further drawbacks mentioned above), they admit that the corrections proposed by their team of one native and one non-native speaker are to be viewed as one possible correct form, not excluding other possible (and plausible) correct forms. The same limitation applies to the present study (also a team of one native informant and the non-native author of the present study) – the corrections inserted by the native informant (speaker of American English) are to be viewed as one possible version out of several target hypotheses.

Further, the difficulty of correcting errors goes back to the concept of error itself and its somewhat elusive definition as "a linguistic form […] which, in the same context […] would, in all **likelihood**, not be produced by the speakers' native speaker counterparts" (Lennon 1991:182, my emphasis). Since native speakers often present several competing target hypotheses about a single learner utterance (cf. Lüdeling 2010), the error annotation by a single native informant is always biased towards this annotator's implicit target norm. The difficulty of error identification is also reinforced by the widened error concept mentioned by Eriksson, "which [includes] an in-between category […] which is nevertheless not obviously nativelike" (Eriksson 2008: 109). Thus, even though the corrected versions and the errors themselves are easily searchable in an error-annotated corpus, the restrictions proposed by Dagneaux et al. (1998) still apply.

This being said, one great asset of the error-detection and error-annotation method used in the present study is the fact that the native informant read the essays very carefully; she was thus able to judge aspect errors within the greater temporal context of each learner

essay and not in isolation. Therefore, the qualitative analysis focuses on both discourse and grammatical accuracy of progressive and perfect forms in Bulgarian and German EFL writing and can thus wonderfully supplement the quantitative findings concerning the under- and overuse of aspect forms in Bulgarian and German EFL writing outlined in the previous two chapters. Lastly, annotating the errors in context is a powerful EA tool, since instances of non-use or avoidance of certain forms can also be captured and measured against the whole number of instances of learner use of the progressive and the perfect – both targetlike and non-targetlike, as well as compared against the number of verbs and words in the learner corpora.

8.2. Classifying Misuse of the Progressive and the Perfect

The next two sections deal with the classification of instances of Bulgarian and German EFL learners' misuse of the progressive and the perfect in terms of two different (surface) misuse types: 1) misuse involving overgeneralisation of the forms in the context of other temporal forms, and 2) misuse involving undergeneralisation or replacement of aspect forms by other tense-aspect forms. 'Overgeneralisation' and 'undergeneralisation' are used here to denote instances of misuse which either involve the **encroachment of one aspect** form in the contexts of other tense-aspect forms or **the non-use or sheer avoidance** of this form respectively; they do not automatically mean that the sources of these errors are purely intralingual and based on e.g. false analogy – it may well be the case that L1 transfer effects and further factors are also at play. A more detailed discussion concerning the identification of possible transfer effects from L1 Bulgarian and L1 German into English and their explanation will be offered in chapter 9.

8.3. Learner Misuse of the Progressive Aspect

The aim of the present section is to describe instances of Bulgarian and German EFL learners' misuse of the progressive aspect in terms of two different types of misuse: 1) misuse involving overgeneralisation of the progressive to non-progressive contexts, i.e. erroneous uses of the progressive (i.e. "traditional" errors of the progressive in e.g. simple present contexts), and 2) misuse involving undergeneralisation or non-use of the progressive in required progressive contexts and its replacement by other tense-aspect forms (e.g. replacement of the progressive by the simple present or the simple past). The progressive has

been identified as a traditional stumbling block for EFL learners by a number of previous corpus-based and learner corpus studies (cf. Virtanen 1997; Axelsson and Hahn 2001; Housen 2000; Lenko-Szymanska 2007; Eriksson 2008; Wulff and Römer 2009; Hundt and Vogel 2011 etc., see also chapter 4), all of which have shown that even advanced EFL learners after many years of instruction and exposure to English still deviate from the native corpus-based norm in various ways, mostly in quantitative terms such as e.g. overuse; however, few learner corpus studies so far have undertaken a detailed qualitative investigation of the types of deviations concerning the progressive in writing and the reasons behind them (e.g. Eriksson 2008 being one notable exception). Since the progressive is a new grammatical category for both Bulgarian and German EFL learners (see chapter 2), a detailed qualitative comparison between the types of misuse of the progressive in advanced Bulgarian and German EFL writing will provide new insights into the way L1-induced differences affect the target-like use of the progressive, as well as into the interplay between other learner- and learning-related factors such as the amount of exposure to English, the proficiency and writing expertise of the learners, as well as the effects of classroom instruction. In addition, section 8.6 will offer new corpus-based methods of learner misuse quantification in accordance with one of the traditional approaches to the measurement of learner accuracy employed in second language acquisition research (cf. Pica 1983).

8.3.1. Overgeneralisation of the Progressive Aspect

A number of learner corpus studies performed so far have shown that the progressive is a major challenge even for advanced EFL learners, who often fail to use it in target-like contexts, and "particularly for learners whose L1 does not have a direct counterpart to the progressive (Wulff and Römer 2009: 116). Apart from the various studies which have identified that EFL learners from a wide range of mother-tongue backgrounds (e.g. Polish, Swedish, German and Finnish) generally overuse the progressive in writing in comparison to native speakers (e.g. (Virtanen 1997; Axelsson and Hahn 2001; Lenko-Szymánska 2007 etc., see chapter 4), a few studies also comment on EFL learners' tendency to extend the progressive onto non-progressive verbs (i.e. stative verbs), as well as non-progressive contexts (e.g. Housen 2002a; Housen 2002b; Eriksson 2008; Hundt and Vogel 2011 etc.). The present section deals with the instances of overgeneralisation of the progressive to non-progressive contexts (i.e. erroneous uses of the progressive in traditional terms) in the writing

of advanced Bulgarian and German EFL learners. Since the progressive as a grammatical category is absent in both German and Bulgarian as native languages, where progressivity is realised via other morphological and lexical means, non-targetlike use of the English progressive in Bulgarian and German EFL writing is likely to accompany its general underuse (see chapter x) and needs further categorisation and clarification. Altogether, 22 instances of overgeneralisation of the progressive in non-progressive contexts and verbs were found in *BUCLE*_110,000 and 57 in *GICLE*_110,000, the overwhelming majority of which (over 95% in *BUCLE*_110,000 and 85% in *GICLE*_110,000) with present time orientation and in contexts where the simple present would have been the preferred form[61].

Replicating the analysis of the general overuse of the progressive with stative and atelic verbs in Bulgarian and German EFL learner writing in comparison to novice and expert native writing presented in chapter 7, a number of instances of incorrect overextension of the progressive to stative verbs were identified in the subcorpora *BUCLE*_110,000 and *GICLE*_110,000. These instances fall under four major categories with respect to the types of verbs and verb phrases the progressive combines with, which will be discussed below: 1) verbs of cognition and physical perception (e.g. *think*, *feel, perceive* etc.), 2) transitive verbs with animate subjects (e.g. *have, tolerate, rule*), 3) locatives and intransitives (e.g. *sit, lie, live*) and 4) miscellaneous passive verb phrases. The following examples illustrate unnatural combinations of the progressive with stative verbs signalling cognition, emotion and physical perception:

8.1. You no sooner buy a new product than you **are thinking** about its replacement <*ICLE*-BG-SUN-0003.1>

8.2. Dehumanization in our modern times starts from this early age when the child **is already perceiving** the things around him […]. <*ICLE*-BG-SUN-0069.1>

8.3. Everybody who has read the example about my father, **is** probably **thinking** my father is crazy; […]. <*ICLE*-GE-AUG-0016.3>

8.4. Once you have found your way in a foreign country and in the foreign language as well **you're feeling** strong and sure of yourself and you're going to master more easily embarrassing and unusual situations. <*ICLE*-GE-AUG-0020.1>

---

[61] A detailed discussion of the rates of overgeneralisation and undergeneralisation of the progressive and the perfect proportionate to the total number of progressive and perfect verb phrases in *BUCLE*_110,000 and *GICLE*_110,000 will be given in section 8.6.

The subjects in these four examples (taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively) are all human agents, and the verbs are all stative verbs expressing cognition, perception or emotion; in examples (8.1) and (8.3), the Bulgarian and German authors refer to generic situations where one particular condition (i.e. someone buying a new product or someone reading about the author's father) would induce a particular thought (i.e. replace the product or think that the father is crazy); in these two cases the native informant would have opted for the simple present form in order to emphasize the genericness of the described situation. The other two examples contain emotive verbs and verbs of physical perception which describe the states of children perceiving the surrounding world and students feeling confident in a foreign country using a foreign language; here again, the native informant would have preferred the simple present form over the progressive form. Assuming that the essay authors wanted to focus on the temporariness of the situations described, and used the progressive with stative verbs combined with active human agents following Huddleston and Pullum's (2005: 167) and Biber et al.'s (1999: 473) categories of admissible and most frequent combinations of the progressive with stative verbs, the four cases of overextension of the progressive to stative verbs could be partially justified; however, these four situations do not imply temporariness of the actions, but rather convey genericness, and therefore the use of the progressive sounds awkward and unnatural. The second category of overgeneralisation of the progressive to stative verbs concerns the use of transitive verbs with human subjects, as shown in the following two examples taken from *GICLE*_110,000:

8.5 This strange behaviour is often very annoying for the inhabitants, but **they are tolerating** a great deal, because **they are depending** on the money from the tourism. <*ICLE*-GE-AUG-0081.1>

8.6 It's already the very sound of that single word "relatives" that arouses the most ambivalent and ambiguous feelings in those people who **are actually having** some relation to their relations - and I can't exclude myself in this connection. <*ICLE*-GE-AUG-0033.3>

In example 8.5 the learner describes a habitual and recurring situation in which the inhabitants of an alpine village have to put up with streams of tourists every year; even though it is combined with a human agent, the progressive form of the verb *tolerate* in the first clause of the sentence does not imply any temporariness or a voluntary change of state (e.g. waxing and waning situation after Huddleston and Pullum 2005: 167). It could therefore be classified as an emotional use of the progressive, where the author wants to convey their negative

attitude towards tourists (cf. Mindt 2000: 256). Still, the second progressive form *depending* in the subordinate *wh*-clause in this sentence sounds even more unnatural than the first one, since the dependence on tourists' money is even less of an agentive action because its human subject is not in control, but more of a passive recipient (cf. Huddleston and Pullum 2005: 167). Example 8.6 is likewise non-targetlike and unnatural, since the stative possessive verb *have* is not used in the temporary sense of e.g. *having a chocolate* or *having a bath* with the human agent, but in the sense of as a generic situation that is stable and cannot be controlled such as having relatives. Therefore, these two examples of the progressive illustrate learners' failure to match the appropriate function onto the correct verb form, which in both cases would have been the simple present. In addition to the combinations of the progressive with transitive stative verbs and human subjects, one example involving a non-human subject was found in *GICLE*_110,000:

> 8.7. Animosity and hatred **are ruling** the world. <*ICLE*-GE-DRE-0024.1>

Here the subject is non-human and non-agentive in Huddleston and Pullum's (2005: 167) framework and the propositional content implies a stable situation rather than a temporary dynamic state; *animosity and hatred* cannot be thus in control of the situation. One alternative reading in support of the progressive would again be the attitudinal reading, when the progressive conveys a negative emotional overtone; nevertheless, this example was still found as awkward and unnatural by the native informant, since it was the introductory sentence of a German learner essay on the topic of "Is it worth living on Earth?" The third type of learner overgeneralisation of the progressive to stative verbs involves combinations of the progressive with locative stative verb phrases (e.g. *live*, *lie*, *sit* etc.) as illustrated in the following examples taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively:

> 8.8. The whites are united by not only they rights to be living beings but with the way of life **they are living** and very often by the idea that blacks are inferior. <*ICLE*-BG-SUN-0226.1>

> 8.9. They are born equal but during the time **they are living** they become more and more different from each other, unequal. <*ICLE*-BG-SUN-0226.1>

> 8.10 As mentioned above, reasons for animosity and hatred **are lying** in man itself. <*ICLE*-GE-DRE-0024.1>

8.11. When I am dreaming about past times I see little villages surrounded by dark woods and many children who gather around the fireplace. A very old grandmother **is sitting** in the rocking chair and is telling fairy tales. <*ICLE*-FR-ULG-0002.1>

The former two examples contain the stative verb *live*, which according to Eriksson (2008: 190) is "interesting because it is a verb with stative qualities which is quite commonly used with the progressive aspect". Quirk et al. (1985: 206) define the verb *live* is a stance verb, and both Quirk et al. (1985: 206) and Biber et al. (1999: 474) state that *live* can be readily used with the progressive in order to express a temporary state in authentic language use (e.g. *Hans is living in Barcelona at present*). Eriksson (2008: 190) argues that it is therefore difficult for EFL learners to differentiate between the uses of *live* in its progressive form and in its simple form and identifies a number of examples of *live* in its progressive form in the Swedish corpus of *ICLE* which are not "felicitously [combined] with state verbs if the dynamising features agent activity, waxing and waning situations and temporariness are absent" (Eriksson 2008: 192). A closer look at the two examples from *BUCLE*_110,000 containing *live* in its progressive form shows that although combined with an agentive subject, the situations described can only be interpreted as ongoing and temporary if we assume that the life of a human being (or people in general) is temporary. Still, one argument against such an interpretation is the fact that in both cases the Bulgarian learners argue on behalf of the white population or on behalf of humankind as a whole, which suggests genericness of the situation, rather than temporariness and blocks a progressive reading.

Apart from *live*, Biber et al. (1999: 474) enumerate further stative verbs such as *stay*, *wait*, *sit* and *stand*, which often have a limited duration and can thus occur in the progressive. Examples 8.10 and 8.11 (taken from *GICLE*_110,000) illustrate such progressive uses of *lie* and *sit*. Similar to the Bulgarian learner instances containing the progressive form of *live*, the German learners here describe generic situations – in the first case the learner argues that animosity and hatred are an intrinsic part of the world, whereas in the second case the learner describes a general condition that whenever they dream, they see an old grandmother sitting in a rocking chair. While the progressive in the second example contributes to a certain immediacy and vividness of the description and such extended uses of the progressives have also been identified in spoken British English by more recent corpus-based studies (e.g. Römer 2005: 95), the native informant in the present study would still have selected the simple form over the progressive form in the context of the whole sentence in both examples.

The fourth type of overgeneralisation of the progressive to stative verb phrases concerns *BUCLE*_110,000 in particular, where the progressive has been extended to miscellaneous passive verb phrases involving achievement and accomplishment verbs in the passive, as illustrated in the following four examples:

8.12. If, on the other hand, we take an inside look at the way study **is being organised**, we will soon notice a certain formula shared between students and lecturers. <*ICLE*-BG-SUN-0071.1>

8.13. And though that **is being** continuously officially **stated** and **re-stated** often the talk about equality remains just a euphemism to hide the cruel reality. <*ICLE*-BG-SUN-0233.1>

8.14. Nowadays everything **is being ascribed** financial value. <*ICLE*-BG-SUN-0233.1>

8.15. The pilgrimage to intellect, in addition, **is being encouraged** by certain modern trends in literature and other kinds of fiction. <*ICLE*-BG-SUN-0011.1>

The verbs *organise*, *state*, *ascribe* and *encourage* are all telic verbs which are of limited duration and have a clear endpoint; however, their respective passive forms *are organised*, *are stated*, *are ascribed* and *are encouraged* are more difficult to interpret in terms of their inherent lexical aspect since they express the resulting state of a situation, rather than the situation itself (cf. Eriksson 2008: 185, see also chapter 7). Example 8.13 and example 8.14 are temporally modified by the adverbials *nowadays* and *continuously* – two adverbials which convey a certain ongoingness of the action and which are often found to modify the progressive; still, their use does not contribute to a greater dynamicity in these two cases, but rather implies a habitual repetitiveness of the described actions. In addition to the lack of dynamicity in the above situations, the subjects in all four examples are all non-human and non-agentive – an additional feature which does not facilitate the use of the progressive and which makes them sound awkward and unnatural. In comparison, only one progressive passive example involving a human subject was found in *GICLE*_110,000:

8.16 There one **is being served** everything from traditional german dishes with thick gravies and dumplings to greek, turkish, italian or chinese foods to more exotic varieties such as japanese sushi, indian rice, or arabian couscous. <*ICLE*-GE-AUG-0052.1>

In this example the German learner describes a generally valid situation concerning the variety of restaurants in Augsburg – a description which tends to be a stable, non-

temporary state and which sounds odd in the progressive; here again, the native informant would have opted for the simple form.

Apart from the instances of overgeneralisation of the present progressive to stative verbs, a number of instances of overgeneralised progressives to dynamic atelic and telic verbs in non-progressive contexts were identified in *BUCLE*_110,000 and *GICLE*_110,000 in particular. The following two examples (taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively) demonstrate the overgeneralisation of the progressive with atelic (activity) verbs:

> 8.17. I **am Daydreaming** and still I believe this is what makes me a human being. <*ICLE*-BG-SUN-0220.1>

> 8.18. Every summer there are million people on their way to their holidays; they **are going** by train, car or aeroplane and travelling all around the world. <*ICLE*-GE-AUG-0081.1>

The use of the progressive in example 8.17 involving the activity verb *daydream* is difficult to categorise, since it can be interpreted in two different ways: as a description of a concrete, ongoing situation of a limited duration (i.e. the Bulgarian learner is daydreaming at one particular moment in the present), or as a habitual/generic situation, where the general act of daydreaming makes the author a human being – the native informant opted for the second interpretation and the respective use of the simple present instead of the present progressive form. Likewise, example (8.18) involving the activity verb *go* implies a habitual situation which is customary to a group of people and repeats itself every year (i.e. people travelling by train, plane or car), and which sounds odd in the progressive despite the active human subject who is in control of the situation. An alternative interpretation of this second case of overgeneralisation of the progressive could be the attitudinal use of the progressive (cf. Mindt 2000: 256), i.e. the progressive functioning as means of expression of the German learner's dislike of the act of travelling; however, judging by the surrounding context, the native informant would have preferred the simple form over the progressive form here as well. In addition to the overgeneralisation of the progressive to habitual and generic situations involving atelic dynamic verbs, a number of overgeneralised progressives with telic verbs such as accomplishments and achievements were found in both learner subcorpora:

8.19. I think a nice solution to this "crisis" in higher education **is coming** our way in the form of the project for a distinction between Bachelor and Magister degrees […]. *<ICLE*-BG-SUN-0068.1>

8.20. The credibility of television seems to be unshakeable, especially when the pictures **are touching** us emotionally. *<ICLE*-FR-ULG-0007.2>

8.21. Cold sweat breaks out me and my heart **is leaping and jumping** like a tennis ball. *<ICLE*-GE-AUG-0053.1>

8.22. An idea **is coming** into my mind. *<ICLE*-GE-AUG-0023.1>

All four verbs (*come*, *touch*, *leap* and *jump*) in the above examples are achievement verbs which are instantaneous and have no duration – combined with the progressive, such achievement verbs either render an iterative interpretation of the described situation (cf. Brinton 1988: 41) or "focus on the process [directly] before the actual event takes place" (Eriksson 2008: 33; 201). The Bulgarian learner in the first example describes a situation which is happening at the moment of speaking (i.e. B.A. and M.A. degrees replacing the old degrees) – thus, an iterative interpretation of this event is not very likely; more plausible seems the author's attempt to focus on the process of the study reform. However, the progressive form of *come* is part of a relative clause with the inanimate subject *solution*, which implies neither repetitiveness nor a controlled process (in contrast, Eriksson's (2008: 33) "Mr K is reaching the top" has a human subject who is in direct control of the process of reaching) – therefore, the native informant identified this progressive as unnatural. The same explanation applies to example 8.22 from the German learner subcorpus – here the inanimate subject *idea* cannot be in active control of the process and a repetitive interpretation seems irrelevant in the instantaneous situation – both factors do not facilitate the use of the progressive. In contrast, examples 8.20 and 8.21 render an iterative interpretation of the actions of *touching* and *leaping and jumping* – in the first case the author refers to the habitual action of watching television and especially emotionally touching pictures, whereas in the second case the author's heart is pounding hard at a particular moment of a limited duration – both situations could possibly be interpreted as repetitive. Still, the progressive in example 8.20 implies a repetitive physical action which sounds awkward and unnatural in the context of the proposition of emotional "touching" of the viewers[62] – therefore, the native informant would have selected the simple present form over the present progressive form. Although the situation in example 8.21 reads better in the progressive due to the inherent repetitiveness of a

---

[62] According to the native informant the simple version still sounds unnatural from a lexical point of view and is possibly due to transfer from German, see also chapter 9

heart beating, the first verb form in this sentence (*breaks out*) is a simple present form which sets up a simple present temporal frame throughout the whole compound sentence – the two progressive verb forms *leaping* and *jumping* in the second clause of this sentence are thus unmotivated and disrupt the clause parallelism that could have been created if all three verbs were in the simple present.

Although the overwhelming majority of the overgeneralised progressives are with present-time orientation, several instances of overgeneralised past progressives (9 in *GICLE*_110,000 altogether and only one in *BUCLE*_110,000) were identified in the two learner subcorpora – all of them in the context of the simple past. The following examples illustrate the overgeneralisation of the past progressive with dynamic verbs to non-progressive past contexts:

> 8.23. Industries **were flourishing**, technology **was developing**, and people **were optimistically looking** towards a good future. <*ICLE*-FR-ULG-0018.2>
> 8.24. Is this what we **were dreaming** when we were kids? <*ICLE*-BG-SUN-0136.1>

In these two examples, the learners describe two dynamic situations in the past which can be classified as ongoing (example 8.24) or waxing and waning (example 8.23) and thus facilitating the use of the progressive; both feature agentive subjects and dynamic verbs; still, the native informant would have opted for the simple past form instead of the past progressive. A closer look at the surrounding linguistic and discourse context of the two examples (sub-examples 8.23.1 and 8.24.1) shows that both learners start developing an argument in the simple present, shift to the simple past and incorporate the past progressive with no apparent reason or justification for its use:

> 8.23.1 Our society is not as stable as it has been a generation ago. After the two wars there was an upward movement throughout Europe. Industries **were flourishing**, technology **was developing**, and people **were optimistically** looking towards a good future. But nowadays optimism is a mentality completely strange to the biggest part of the younger generation. . <*ICLE*-FR-ULG-0018.2>

> 8.24.1 The whole system operating in the modern world seems meaningless or maybe its meaning leaves little place for dreaming and imagining. It is like we are in a vicious circle - we study hard so that we could get a proper job, work hard so that we could lead a proper life and send our children to proper schools where they will study hard so that they could get a proper job and so on and so on . Is this what **we were dreaming** when we were kids? I think not. <*ICLE*-BG-SUN-0136.1>

Analysed in isolation, these two examples of overgeneralisation of the past progressive are not necessarily wrong; however, based on the larger discourse context of the learner essays, it becomes obvious that similar to advanced Swedish EFL learners, Bulgarian and German EFL learners are also inconsistent in their use of tense and aspect and fail "to adopt a clear discourse strategy" (Eriksson 2008: 210), thus making unmotivated shifts between non-progressive and progressive forms. Even if the progressive would otherwise sound plausible in the above examples, the wider linguistic and discourse context does not support its use – a special emphasis on the act of dreaming or the development of industrialised societies as temporary and ongoing actions of limited duration is thus unmotivated and unnecessary.

To summarise, the present section focused on the instances of overgeneralisation of the present and past progressive to non-progressive contexts in advanced Bulgarian and German EFL writing. Three major types of advanced learners' misuse of the progressive in non-progressive contexts were identified:

(1) **combinations** of the progressive with various stative verbs
(2) **overgeneralisation** of the progressive to generic and habitual situations
(3) **discourse-dependent misuse** of the progressive

Since the overgeneralisation or misuse of the progressive in non-progressive contexts is not the only type of misuse identified in the two learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000, the following section will deal with instances of non-use of the progressive aspect (avoidance) and its replacement by other tense-aspect forms.

8.3.2. Undergeneralisation of the Progressive aspect

The comprehensive error-tagging of all verb phrases in the subcorpora *BUCLE*_110,000 and *GICLE*_110,000 not only lends itself to "traditional" corpus-based error analyses of learner overgeneralisation of the progressive to non-progressive contexts, but it also allows for an investigation of learner non-use of the progressive in required progressive contexts. This section examines the undergeneralisation or non-use of the progressive by looking at the contexts where the progressive should have been used within the temporal and discourse

context according to the native informant, but was replaced by its simple aspect counterpart instead. A closer examination of the occurrences of undergeneralisation of the progressive in the error-tagged subcorpora shows that in the overwhelming majority of the cases, the simple present has been used instead of the present progressive, followed (to a much lesser extent) by the replacement of the past progressive by the simple past. In a similar vein, Eriksson (2008: 207) found several examples of non-use of the present progressive and its replacement by the simple present in the Swedish component of *ICLE* and noted that although rare, this type of misuse was still present in advanced Swedish learners' written English. In the present study, 33 instances of undergeneralised progressives altogether were found in *BUCLE*_110,000 and 50 examples in *GICLE*_110,000. In examples 8.25 and 8.26 taken from *BUCLE*_110,000, the simple present has been used to describe dynamic and ongoing situations which are incomplete and still subject to change:

> 8.25. Finally, this regards the culmination of human-technological genius – the computer which **becomes** more and more widely-used by millions of people. <*ICLE*-BG-SUN-0250.1>

> 8.26. And unfortunately problems such as racism and class division **grow** bigger and bigger. <*ICLE*-BG-SUN-0226.1>

In the former example, the action of becoming widely used is adverbially modified by *more and more*, an adverbial which according to Huddleston and Pullum (2005: 167) refers to waxing and waning situations when modifying stative verbs and which facilitates the use of the progressive with non-stative verbs too (cf. also Eriksson 2008: 209). The latter example is similar insofar as it conveys an ongoing, changing situation (racism growing bigger and bigger) and contains a comparative use of the adjective *big*, which also presupposes the use of the progressive rather than the simple present. The use of the simple present conveys a habitual interpretation in both cases and is thus inconsistent with the actual dynamicity of the ongoing situations. Similar uses of the simple present instead of the present progressive were found in *GICLE*_110,000 too:

> 8.27. Many historical building **fall victim** to the growing pollution. <*ICLE*-GE-AUG-0011.2>

> 8.28. They **lose** more and more the image of a simple, naive house-wife, because they get another perspective and another shaping as members of the working-class. <*ICLE*-GE-AUG-0062.1>

Again, these two examples refer to ongoing, dynamic situations which are undergoing continuous changes at the present moment – in the first case buildings are increasingly falling prey to the pollution, and in the second case modern women are losing the image of traditional housewives, again adverbially modified by *more and more*. The native informant would have opted for the use of the present progressive in all four cases; the Bulgarian and German EFL learners have opted for the much less suitable and in all four cases rather unnatural-sounding simple present form. Further occurrences of the non-use of the progressive in progressive contexts and its replacement by the simple present concern temporary situations in progress, which happen at a given, limited time, as shown in the following examples:

8.29. A bus of people of all shades of colour **travels** through the desert in South Africa. *<ICLE*-BG-SUN-0255.1>

8.30. We either **run** away from something, or search inspiration and ideas. *<ICLE*-BG-SUN-0222.1>

8.31. She comes on Fridays at 2 o'clock p.m., when most of the people working in the bureau have already left or still **buzz** around like workaholics-depending on their average earning. *<ICLE*-GE-AUG-0002.2>

These three examples (taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively) depict classic temporary situations in progress which, however, are rendered habitual via the use of the simple present and thus sound awkward and unnatural. In line with Eriksson's (2008: 208) findings that the absence of a temporal adverbial more often than not triggers non-use of the progressive, the majority of the occurrences of non-use of the present progressive in progressive contexts and its replacement by the simple present happen when there is no temporal specification signalling a present, ongoing situation; still, there is one example in *GICLE*_110,000 and several examples in *BUCLE*_110,000, where the simple present is modified by the temporal adverbials *now* and *nowadays*:

8.32. **Nowadays** the formerly communist Easteuropean countries **tend to** adopt the capitalistic and democratic system of Western nations. *<ICLE*-FR-ULG-0012.2>

8.33. To some extent, that is what we **experience now**, three centuries later. *<ICLE*-BG-SUN-0070.1>

8.34. **Nowadays** when **industrialization enjoys** its zenith, more and more people are worshipping their new idols-all the fruits of scientific progress, which are supposed to make their life easier. *<ICLE*-BG-SUN-0220.1>

8.35. **Now we dream** how to invent a new kind of light bulb that will not need changing every other month […]. <*ICLE*-BG-SUN-0049.1>

The above four examples illustrate four different situations in progress modified by *now* and *nowadays*, which are thus concurrent with the moment of speaking; still, the learners have opted for the simple present instead of the present progressive, which would have been the preferred option. Two other temporal adverbials signalling an ongoing moment and modifying the simple present instead of the present progressive were found in *BUCLE*_110,000:

8.36. **At the moment** I **teach** two friends of mine English and I can say without exaggerating at all that I have no problems with practising the profession of a teacher. <*ICLE*-BG-SUN-0109.1>

8.37. New theories are too, **constantly developed**; research provides with new and interesting insights in the concerned sphere, thus founding the basis for scientific progress. <*ICLE*-BG-SUN-0064.1>

Again, *at the moment* and *constantly* convey a dynamic, ongoing situation of teaching one's friends (example 8.36) or developing theories (example 8.37); however, the Bulgarian learners have opted for the non-progressive form which conveys a rather stative, habitual meaning and which sounded awkward and unnatural to the native informant who conducted the error-tagging. In addition to the non-use of the present progressive and its replacement by the simple present, a number of examples of non-use of the past progressive and its replacement by the simple past were found in *GICLE*_110,000[63]:

8.38. He **did not refer** to heavy metal or trash metal, which is in fact a very aggressive style of music, but **he referred** to Rock music in general. <*ICLE*-GE-SAL-0031.5>

8.39. It was one of these cold and grey winterdays in December, a few days before Christmas when **I hurried** through the city to get at least a few of the mountains of christmas presents I still had to buy. <*ICLE*-GE-AUG-0097.1>

8.40. When I **leafed through** a weekly magazine recently, my eyes were magically attracted by the photo of a little black girl of about three. <*ICLE*-GE-AUG-0014.3>

In these examples, the native informant would have chosen the past progressive over the simple past due to the temporariness and limited duration of the situations described in

---

[63] No occurrences of the replacement of the past progressive by the simple past were found in *BUCLE*_110,000

examples (example 8.38) and (example 8.39) – referring to a particular kind of music and hurrying through the city at a particular point of time in the past, as well as due to the temporal frame of leafing through a magazine which includes the act of seeing a photo in this magazine (cf. Quirk et al. 1985: 209). Over 30% of all instances of undergeneralisation of the progressive in *GICLE*_110,000 are due to the non-use of the past progressive and its replacement by the simple past (no instances of confusion between the past progressive and the simple past were found in *BUCLE*_110,000): this finding goes hand in hand with the overall distribution of past forms in *GICLE* (see chapter 6), which are significantly more frequent than the progressive and perfect forms with past orientation in *BUCLE*.

Even though functional undergeneralisation of the progressive is difficult to pinpoint and becomes only obvious within the larger context of the learner essays, the above examples of non-use of the progressive in traditional progressive contexts signalled by the ongoing actions and temporal frames described by the verbs, or by the time adverbials accompanying them indicate that Bulgarian and German EFL learners have not yet fully mastered the functions of the progressive aspect, even at an advanced proficiency level and after many years of EFL instructions. The identification of an almost equal number of instances of undergeneralisation and overgeneralisation of the progressive in *GICLE*_110,000 (50 undergeneralised forms vs. 57 overgeneralised forms) and an even greater number of undergeneralised progressives in *BUCLE*_110,000 (33 undergeneralised forms vs. 23 overgeneralised forms) is remarkable insofar as it reveals that both learner populations systematically fail to match the appropriate functions onto the correct tense-aspect forms in EFL writing. The next section deals with instances of learner misuse of the perfect aspect, starting with overgeneralisation of the perfect to non-perfect contexts and finishing the discussion with undergeneralisation of the perfect aspect and its replacement by the simple present and simple past. The last two sections in the present chapter offer a method of learner misuse quantification, as well as a discussion and an explanation of the factors behind advanced EFL learners' misuse.

8.4. Learner Misuse of the Perfect Aspect

Very few learner corpus studies so far have focused on intermediate and advanced EFL learners' misuse of the perfect aspect and the present perfect in particular, Eriksson (2008) and Davydova (2011) being two notable exceptions. On the basis of the Swedish component

of *ICLE*, Eriksson (2008: 139-143) found that advanced Swedish learners misused the present perfect in the context of other tense-aspect forms, mostly due to learners' confusion of the present passive which emphasises the result of a situation and the present perfect which refers to the process leading up to a state (cf. Eriksson 2008: 139). Contrary to his expectations and the contrastive differences in the use of the present perfect in English and in Swedish, instances of an overt misuse of the present perfect in simple past contexts were found to be relatively rare (cf. Eriksson 2008: 141). Along similar lines, Davydova (2011: 270 – 271; 288) noted that due to its semantic complexity, "the present perfect is a challenge to any non-native speaker" (Davydova 2011: 270) and found for her sample of intermediate Russian and German EFL learners that it "tends to be undergeneralised in present perfect contexts and rarely overgeneralised to definite past time reference contexts" (Davydova 2011: 271). Similar to the discussion of advanced Bulgarian and German EFL learners' misuse of the progressive presented in the previous section, the aim of the present section is to identify and discuss instances of over- and undergeneralisation of the perfect and in particular the present perfect for these two learner populations – and ultimately to find a common denominator between the previous learner corpus studies and the present study with regard to the types of learner misuse of the perfect.

8.4.1. Overgeneralisation of the Perfect Aspect in Non-Perfect Contexts

The first type of learner misuse of the perfect aspect concerns the overgeneralisation of the perfect to non-perfect contexts, i.e. the use of the perfect instead the simple present or the simple past in *BUCLE*_110,000 and *GICLE*_110,000. Altogether, 64 instances of overgeneralisation of the perfect were found in *BUCLE*_110,000 and 59 in *GICLE*_110,000; the great majority of these instances were found to be due to overgeneralisation of the present perfect to non-present-perfect contexts (61% of all instances of overgeneralisation in *GICLE*_110,000 and 78% in *BUCLE*_110,000). In most cases, the present perfect was used by the Bulgarian and German learners instead of the simple past, and to a much lesser extent instead of the simple present. The following four examples taken from *BUCLE*_100,000 and *GICLE*_110,000 respectively illustrate 'classic' instances of learner misuse of the present perfect in simple past contexts indicated by a past-time adverbial such as *ago* or *yesterday*:

8.41. **I've only found** the answer a few days **ago**. *<ICLE*-BG-SUN-0027.1>

8.42. It is a truth generally acknowledged that progress sometimes exceeds the limited capacity of human imagination of people who **have lived** long **ago**. *<ICLE*-BG-SUN-0043.1>

8.43. Our society is not as stable as it **has been** a generation **ago**. *<ICLE*-FR-ULG-0018.2>

8.44. **Have you been yesterday** on the report on the growth of the rubbish-mountains? *<ICLE*-GE-AUG-0033.1>

These examples show that both learner groups have difficulties differentiating between the present perfect and the simple past and are equally insensitive to temporal contexts signalled by past-time adverbials which clearly refer to a moment in the past that has come to an end (cf. Lim 2007). Example 8.41 could also be interpreted as a direct negative transfer from German, where the respective German translation would have been perfectly acceptable due to the German *Perfekt* subsuming most of the functions of the German preterite and occurring freely in combinations with past-time adverbials in narrative contexts. Still, most instances of overgeneralisation of the present perfect to non-present perfect contexts are temporally unspecified and become only clear in the larger temporal and discourse context of the essays:

8.45. Every revolutionary invention **has first become** alive in man's dreams. *<ICLE*-BG-SUN-0055.1>

8.46. The historic handshake between Mr. Bush and Mr. Gorbatchov on the lawn of the White House **has put an end** to the division of the world that had lasted for almost half a century. *<ICLE*-GE-AUG-0025.3>

Here the choice between the present perfect and the simple past is slightly more complicated, especially in example 8.45, where the whole sentence does not specify a completed action itself (i.e. men may still keep on inventing in their dreams), but where the surrounding linguistic context points to a completed action similar to Elsness' (1997: 37) example that a sentence like "Shakespeare has written some of the most beautiful poetry" is only acceptable if the topic is poetry and not Shakespeare who is long dead and no longer writing. Along similar lines, Bush and Gorbatchov's handshake after the fall of the Berlin wall is a completed action that has long come to an end and that requires the simple past, rather than the present perfect as a linguistic option. Other ungrammatical uses of the present perfect in simple past contexts appear to indicate an inability on the part of the learners to

recognise the temporal continuity of a narration created by the simple past, as illustrated in examples 8.47 and 8.48:

> 8.47. Many eras **have passed** till man **discovered** simple truths as the fact that the earth is round, for instance, and those mentioned above (*ICLE*-BG-SUN-0028.1)

> 8.48. On the other hand it **has also been** one of the grey little boxes that **broke** the news about my sisters new born baby-boy (*ICLE*-GE-AUG-0023.1)

In both sentences, the native choice would not have been the present perfect, since the events they refer to are clearly situated in the past, with an obvious gap between their completion and the present moment (e.g. Quirk et al., 1985: 183) which is signalled by the rest of the sentence related in the simple past. However, the learners here seem to have confused the past perfect form with the present perfect form in their desire to pinpoint an event which happened prior to the time referred to by the simple past in the rest of the sentence. The American informant would still have opted for the simple past tense in order to signal a sequence of happenings in the past, rather than the 'past-in-the-past'-signalling past perfect (Quirk et al., 1985: 195). Such temporally unspecified instances of misuse of the present perfect in non-present-perfect contexts are often difficult to analyse, since many of them are interclausal, meaning that the misuse cannot be easily identified in isolation, but the larger context needs to be consulted (cf. Eriksson 2008: 125). Much less frequent are instances of overgeneralisation of the present perfect to simple present contexts (12 examples in *GICLE*_110,000 and 7 in *BUCLE*_110,000 altogether), as demonstrated in the following examples:

> 8.49. When imagination **has exceeded** the technical power of its times, its products can later be constructed, its fantasy world can be recreated, but this time in reality. <*ICLE*-BG-SUN-0043.1>

> 8.50. Everybody **who has read** the example about my father, is probably thinking my father is crazy; […]. <*ICLE*-GE-AUG-0016.3>

Here the native informant would have preferred the simple present instead of the present perfect due to the general validity of the actions described by the verbs *exceed* and *read*, which can be assumed to signify real conditions – i.e. the condition of imagination exceeding technical power and the reader reading the example about the German author's father. The present perfect sounds in these examples unnatural insofar as the two actions are

clearly situated in the present, rather than the recent past, which is also signalled by the surrounding context; thus, a 'current relevance' reading is hardly possible.

Last, but not least, over 35% (21 instances) of the overgeneralisation of the perfect in inappropriate contexts in *GICLE*_110,000 was found to be due to overgeneralisation of the past perfect to non-past-perfect contexts; in contrast, only 18% or 12 cases of overgeneralised past perfects were found in *BUCLE*_110,000 altogether. This finding corroborates yet again the general observation that progressive and perfect forms with past-time orientation are a lot more frequent in *GICLE* than in *BUCLE*; the subcorpora *BUCLE*_110,000 and *GICLE*_110,000 are no exception to the rule (see chapter 6). Overgeneralisation of the past perfect occurs almost exclusively in simple past contexts in both subcorpora, as demonstrated by the next two examples:

> 8.51. Lets take for instance the famous mathematician and physicist, Sir Isaak Newton who **had discovered** the laws of mechanics and gravitation. <*ICLE*-BG-SUN-0028.1>

> 8.52. Although she **had graduated** in communication studies, it was impossible for her to get a job in New Zealand. <*ICLE*-GE-AUG-0006.1>

In both cases, the native informant would have preferred the simple past form to the past perfect form; however, the Bulgarian and the German learners in these two examples have opted for the past perfect form in an attempt to relate a situation which happened prior to the events related by the simple past in the main story line; nevertheless, such a shift into the pre-past sounds unnatural (in particular in example 8.51) and would not have been necessary in both cases. Other miscellaneous instances of misuse of the past perfect in simple past contexts in *GICLE*_110,000 concern inappropriate uses of the past perfect as part of indirect speech clauses (see example 8.53) or as part of main clauses which do not convey information referring to pre-past events (example 8.54):

> 8.53. Her father is red as a beetroot - if she **had forgotten** that they wanted to spend some days out in the green - he yells at her. <*ICLE*-GE-AUG-0079.1>

> 8.54. The whole spectacle **had been** an enormous success, and I'm sure that it was a great cultural enrichment for the town and even the whole land. <*ICLE*-GE-AUG-0073.1>

In these two examples, the past perfect is incorporated in a narrative story line predominantly carried by the simple present where past forms appear unnecessary, if not

unwanted, and tend to disrupt the tense continuity created by the simple present, thus confirming Eriksson's (2008) and Granger's (1999) results with respect to advanced EFL learners' unmotivated tense shifts in writing. On the whole, overgeneralisation of the past perfect to non-past-perfect contexts seems to be more problematic for German EFL learners than for Bulgarian EFL learners and goes hand in hand with the general quantitative tendency for German EFL learners to prefer aspect forms with past orientation to a greater extent than Bulgarian EFL learners. This tendency will be further discussed in the light of transfer effects from the native language, as well as the role of writing expertise in the final chapter. The following section will give an overview of the types of undergeneralisation or non-use of the perfect aspect by advanced Bulgarian and German EFL learners.

8.4.2. Undergeneralisation of the Perfect Aspect

Similar to the discussion of the undergeneralisation of the progressive, this section examines undergeneralisation or non-use of the perfect by looking at the contexts where the perfect should have been used according to the native informant, but was replaced by the simple past or the simple present instead. Altogether, 93 instances of undergeneralised perfects were found in *GICLE*_110,000 and 59 instances in *BUCLE*_110,000; the overwhelming majority of the cases where the perfect should have been used were learner uses of the simple past (77% of the undergeneralised perfect forms in *GICLE*_110,000 and 62% in *BUCLE*_110,000), followed by uses of the simple present and occasionally by uses of other tense-aspect forms. Similar to the undergeneralisation of the progressive, undergeneralisation of the perfect aspect affects mainly the non-use of the present perfect in present perfect contexts (82% of the undergeneralised perfects in *GICLE*_110,000 and 95% should have been present perfect forms). The following four examples taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively illustrate advanced Bulgarian and German EFL learners' non-use of the present perfect in contexts where the present perfect would have been the preferred form due to the resultative situation described in the sentences, which still pertains to the present moment:

> 8.55. The progress **we made**, are making and will make in these areas is due to those who give wings to their imagination and are not afraid of dreams. <*ICLE*-BG-SUN-0065.1>

8.56. The machines **we invented** help us. <*ICLE*-BG-SUN-0101.1>

8.57. The invention of the telephone **changed** our lives fundamentally. <*ICLE*-GE-AUG-0026.1>

8.58. Nobody, neither my mother nor the men I loved, **succeeded** in achieving a change, slight it may be. <*ICLE*-GE-AUG-0050.3>

In the first three examples, the authors talk about the results of past actions or events such as the technological progress of mankind and the inventions of machines like telephones which have changed people's lives using the simple past, even though from an interclausal point of view they keep using the simple present in the surrounding context of the argumentation. Thus, even though not necessarily wrong, the use of the simple past in these examples sounds awkward and unnatural, and the native informant (a speaker of American English who is also expected to prefer the simple past over the present perfect in such contexts, see also chapter 2) would have preferred the present perfect as an indefinite resultative past alternative to the simple past. Although not immediately obvious, the undergeneralisation of the present perfect and its replacement by the simple past in the fourth example becomes again obvious when one looks beyond clause boundaries – the surrounding story line is delivered in the simple present and a sharp shift into the past seems unmotivated. In addition to the temporally unspecified instances of undergeneralisation of the present perfect and its replacement by the simple past, a number of simple past instances modified by classic present perfect temporal adverbials such as *since, recently, lately* and *just* were detected in both learner subcorpora:

8.59. But **recently we began** to realize that something was lost in our education. <*ICLE*-BG-SUN-0095.1>

8.60. **Since then**, people **were sent** to the Space, vaccines against lethal diseases were discovered, bridges were built. <*ICLE*-BG-SUN-0101.1>

8.61. Then suddenly when the murderer **just opened** the door holding a long and shiny knife in his hand creeping carefully […] you suddenly see an average looking woman holding up a box of washing powder. <*ICLE*-GE-AUG-0087.1>

8.62. To cut a long story short, **since then I feel** that cycling not only keeps me in good exercise but also in excellent health. <*ICLE*-GE-AUG-0054.3>

The above instances show that the Bulgarian and German EFL learners from these examples have failed to grasp one of the central meanings of the perfect aspect – the meaning

of anteriority, which is "signalled by tense or by other elements of the sentence or its context" (Quirk et al. 1985: 190), within which the action described by the verb takes place. In the above cases, the actions and events described are framed by a temporal adverbial phrase such as *since + NP* or by a time adverbial referring to the recent past such as *just* and *recently*, all of which are commonly used with the present perfect and sound unnatural, if not wrong with the simple past.

In addition to the replacement of the present perfect by the simple past in temporally framed present perfect contexts, a number of instances of replacement of the present perfect by the simple present, despite the use of temporal adverbials commonly modifying the present perfect were identified in *BUCLE_*110,000 (22 instances) and to a lesser extent in *GICLE_*100,000 (17 instances). The following two examples illustrate such uses:

> 8.63. The fact that foreigners are no longer welcome in Germany **is known** to the public at least **since** the asylum debates at the Bundestag, which were the main issue of German newspapers for a long time. <*ICLE*-GE-AUG-0070.1>

> 8.64. This means that it is not capable of realizing the real change of the world and that the notions of 'good', 'bad', 'beautiful', 'true' or 'untrue' **are** more or less the same **since** Ancient Greece. <*ICLE*-BG-SUN-0092.1>

In these two examples, the action described by the verb in the simple present is temporally framed by an indefinite point in time (the asylum debates in Germany or Ancient Greece), which is signalled via a temporal adverbial phrase; in both cases, the native informant would have opted for the present perfect instead of the simple present. Other instances of non-use of the present perfect and its replacement by the simple present concern learners' confusion between the present passive and the present perfect passive, a problem which has been encountered and discussed by Eriksson on the basis of his sample of advanced Swedish EFL learners (cf. Eriksson 2008: 138 – 139). The following examples illustrate this problem:

> 8.65. It **is allowed** to sell gene-food, for instance manipulated tomatoes, in the U.S.A. since 1991. <*ICLE*-GE-DRE-0009.1>

> 8.66. As far as life outside university is concerned, my expectations **are fulfilled**. <*ICLE*-BG-SUN-0083.1>

Similar to Eriksson's findings, the learners here seem to focus exclusively on the result of the situation expressed by the past participle of the present passive construction, rather than focusing on the "process leading up to the state" (Eriksson 2008: 139) expressed by the present perfect; a focus which seems to be stronger in English than in Bulgarian, German or Swedish as native languages. Temporal modification of the present passive by 'stereotypical' present perfect adverbials and adverbial phrases like since + NP occurs here as well; however, the majority of the uses of the present passive instead of the present perfect passive in *BUCLE*_110,000 and *GICLE*_110,000 were found to be temporally unspecified.

Finally, the native informant identified a number of instances of replacement of the past perfect by other tense-aspect forms, in particular in *GICLE*_110,000 and to a much lesser extent (only 3 instances) in *BUCLE*_110,000. In the overwhelming majority of these cases (almost 100%), the simple past has been used instead of the past perfect, as demonstrated in the following two examples taken from *BUCLE*_110,000 and *GICLE*_110,000 respectively:

8.67. It **was** long ago since she swapped the real world – the veranda. *<ICLE*-BG-SUN-0146.1>

8.68. After three ours of standing more than driving I was nearly as far away from the hospital as **I was** three hours before. *<ICLE*-GE-AUG-0008.2>

The non-use of the past perfect and the misuse of the simple past in example (8.67) becomes obvious when looking at the surrounding temporal and discourse context, as well as due to the combination of the simple past and the temporal adverbial phrase *since + NP*; whereas the misuse of the simple past instead of the past perfect in the second case is even more transparent, since the shift to a pre-past of the main narrative line (i.e. three hours before the state of being far away from the hospital related in the simple past) is not linguistically signalled by the use of a pre-past, i.e. past perfect form. Considerably more frequent (especially in *GICLE*_110,000) are instances of the replacement of the past perfect by the simple past in conditional clauses, as illustrated in examples 8.69 and 8.70:

8.69. I would definitely go back to working out in the gym, **if only there was** one. *<ICLE*-GE-AUG-0010.4>

8.70. Just imagine what could have happened, or to be more precise, what wouldn't have happened if **it wasn't for** this genious' imagination. *<ICLE*-BG-SUN-0028.1>

In these cases, the simple past form is supposed to render an unreal condition – the German author of the former example would go to the gym, but there is no such gym, and the Bulgarian author of the second example ponders on what would have happened without Newton's ingenious inventions which are a historical fact. Conditionals of this type render impossible conditions which require the past perfect form in the *if*-clause to realise the impossibility or irreversibility of the described situation; therefore, the use of the simple past instead of the past perfect seems awkward, if not wrong. In sum, both the overgeneralisation and undergeneralisation of the perfect aspect indicate that both learner groups, advanced as they may be, still experience great difficulties incorporating the perfect appropriately in the discourse and temporal context of their essays, and tend to overgeneralise it to inappropriate contexts or not use it at all. The next section will offer one possible method of quantification of advanced EFL learners' aspect misuse, whereas the final section will comment on the difficulties and possible reasons behind Bulgarian and German EFL learners' misuse in more detail.

8.5. Measuring Bulgarian and German EFL learners' Aspect Misuse: Issues of Misuse Quantification

Having discussed the two major types of learner misuse of the progressive and the perfect in the previous two sections, the present section turns to possible ways of measuring and comparing Bulgarian and German EFL learners' misuse rates in quantitative terms. Since all verb phrases in the subcorpora *BUCLE*_110,000 and *GICLE*_110,000 have been both error- and POS-tagged and the instances of overgeneralisation and undergeneralisation counted, an adapted version of the CEA approach (cf. Dagneaux et al. 1998) was employed in order to quantify learner misuse of the progressive and the perfect. Given that a traditional computer-aided error analysis would have counted the overt errors or overgeneralised progressive and perfect verb forms in proportion to all progressives and perfect verb forms (or the number of words in total) only, disregarding undergeneralisation or learner non-use, the present analysis adopts an adapted approach to learner misuse which accounts for both types of learner misuse dealt with in the previous sections. In order to normalise the instances of over- and undergeneralisation of the progressive and the perfect, the number of finite verb phrases in the two subcorpora *BUCLE*_110,000 and *GICLE*_110,000 was defined first, followed by an extraction of the progressive and perfect verb phrases following the same methodological

procedures outlined in chapter 5. Table 8.1 illustrates the absolute and normalised frequencies (progressive and perfect VPs in proportion to all finite VPs in %) of the two aspect forms for the subcorpora *BUCLE*_110,000 and *GICLE*_110,000.

	progressives total	progressive VPs in %	perfects total	perfect VPs in %
*BUCLE*_110,000	212	1.7	693	5.6
*GICLE*_110,000	339	2.7	540	4.3

Table 8.1. Frequencies of the progressive and the perfect in the subcorpora *BUCLE*_110,000 and *GICLE*_110,000

In terms of the distribution of finite verb forms in *BUCLE*_110,000 and *GICLE*_110,000, the two subcorpora have approximately the same number of finite verb phrases (both in absolute terms and in proportion to all POS-tags), i.e. they are equally 'verby'; however, in terms of the distribution of the two aspect forms, there are highly significant differences concerning *BUCLE*_110,000 learners' quantitative preference for the perfect (p<0.001) and *GICLE*_110,000 learners' quantitative preference for the progressive (p<0.001). In order to normalise and measure the instances of learner overgeneralisation and undergeneralisation of the two aspect forms in proportion to the number of progressive and perfect verb phrases altogether (see table 8.2.), Pica's (1983) adapted version of the target-language use (TLU) analysis quantification method was employed.

	*BUCLE*_110,000	*GICLE*_110,000
overgeneralised progressive VPs	23	57
undergeneralised progressive VPs	33	50
total number progressive VPs	212	339
overgeneralised perfect VPs	64	59
undergeneralised perfect VPs	59	93
total number perfect VPs	693	540
total number of finite VPs	12474	12428
total number of words	112064	113230

Table 8.2. Frequencies of the overgeneralised aspect forms in the subcorpora

Pica's TLU method of morpheme quantification takes into consideration both the incorrect use or non-use of a particular morpheme in required contexts and the over-suppliance or overgeneralisation of this same morpheme in inappropriate contexts (cf. Pica 1983: 70 – 71). The TLU score is calculated with the formula:

$$\text{TLU} = \frac{\text{correct suppliance in obligatory contexts}}{\text{N obligatory contexts + number suppliance in non-obligatory contexts}}$$

Since overt misformations (i.e. incorrect forms) of the perfect and the progressive in advanced Bulgarian and German EFL writing were hardly found (e.g. occasional misformations of the past participle form of the perfect were attested), the TLU formula was adapted to account for the correct suppliance and non-suppliance of the progressive and the perfect in required contexts, as well as for the overgeneralisation of these aspect forms to inappropriate contexts in functional terms. The required contexts in *BUCLE*_110,000 and *GICLE*_110,000 were determined on the basis of the total learner frequency of the progressive and the perfect (see table 8.2) minus the number of overgeneralised uses of the progressive and the perfect in non-required contexts plus the number of undergeneralised instances of non-use of the progressive and the perfect in required contexts. In this way, both undergeneralisation and overgeneralisation could be measured: to illustrate, the percentage of non-use of the progressive or the perfect in required progressive or perfect contexts was calculated by dividing the number of aspect forms which were not supplied in required contexts by the number of all required contexts for these aspect forms:

$$\% \text{ Undergeneralisation}_{\text{progressive/perfect aspect}} = \frac{\text{N non-use in required}_{\text{progressive/perfect}} \text{ contexts} \times 100}{\text{N required}_{\text{progressive/perfect}} \text{ contexts}}$$

Likewise, the over-suppliance or overgeneralisation of the progressive and the perfect in the learner subcorpora was calculated by dividing the number of overgeneralised or over-supplied progressive or perfect verb forms by the number of all other remaining finite verb phrases, where the remaining finite verb phrases were counted by subtracting the number of required contexts for the progressive or the perfect from the total number of finite verb phrases (i.e. all finite verb phrases which are non-progressive or non-perfect):

$$\% \text{ Overgeneralisation}_{\text{progressive/perfect}} = \frac{\text{N over-suppliance}_{\text{progressive/perfect}} \times 100}{\text{N finite verb phrases} - \text{N required contexts}_{\text{progressive/perfect}}}$$

This second measurement allows for a count of the instances of "encroachment" of progressive and perfect verb forms onto the required contexts of other tense-aspect forms in the learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000. Subsequently, both undergeneralisation or non-use of the progressive and the perfect in required contexts, as well as their overgeneralisation to inappropriate contexts were calculated in percentages in order to compare the two learner populations' misuse rates. Figure 8.2 shows the percentages of undergeneralisation or non-use of the progressive and the perfect in the two learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000 calculated with the help of the first formula.
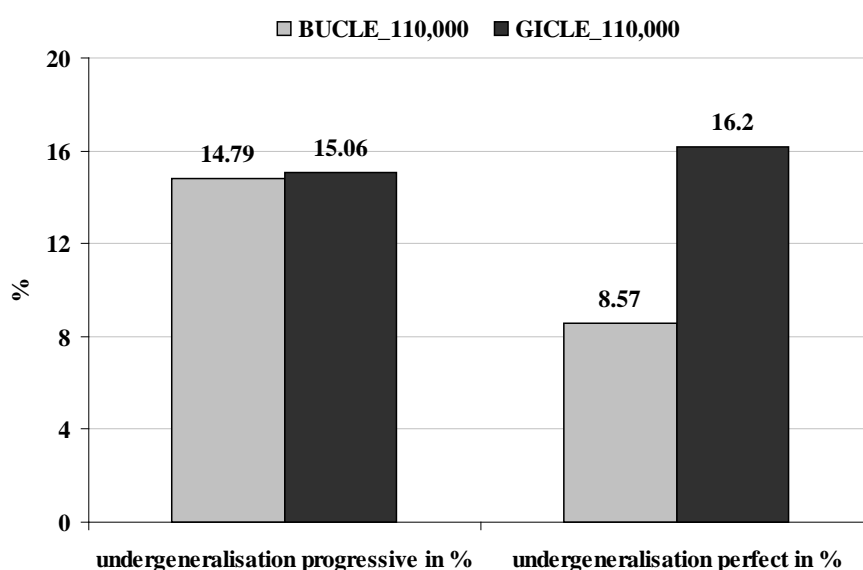
Figure 8.2. Undergeneralisation of the progressive and the perfect in %

On average, the undergeneralisation of the progressive in both learner corpora is higher than the undergeneralisation of the perfect: however, the ratio of undergeneralised perfects in *GICLE*_110,000 is strikingly higher (almost double) than the ratio of undergeneralised perfects in *BUCLE*_110,000 (p<0.0001). Although relatively high, there are no statistically significant differences between the ratios of undergeneralised progressives in the two learner corpora – both learner groups fail to supply the progressive in required contexts equally often. Employing the second type of measurement – the measurement of overgeneralisation of the progressive and the perfect to non-progressive and non-perfect contexts reveals a very different picture: whereas the ratios of undergeneralisation were measured against the total learner frequency of the progressive and the perfect, the ratios of overgeneralisation were measured against the total number of finite verb phrases minus the

required contexts for the progressive and the perfect, thus yielding considerably lower percentage results (figure 8.3).
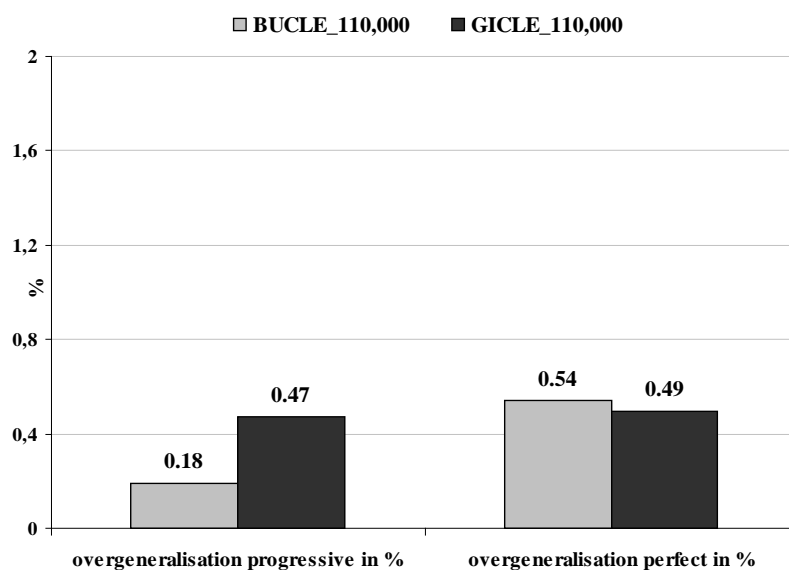
Figure 8.3. Overgeneralisation of the progressive and the perfect in %

The results for the ratios of overgeneralised progressives and perfects supplied in inappropriate contexts show no significant differences between *BUCLE*_110,000 and *GICLE*_110,000 in terms of the overgeneralisation of the perfect, however, they show significant differences in terms of the overgeneralisation of the progressive – *GICLE*_110,000 learners overgeneralise the progressive extending it to non-progressive contexts to a much greater extent than *BUCLE*_110,000 learners ($p<0.001$). Nevertheless, overgeneralisation of the perfect and its suppliance to non-perfect contexts is on average higher than the overgeneralisation of the progressive and its suppliance in inappropriate non-progressive contexts. A word of caution is in order here: the percentage rates of overgeneralisation and undergeneralisation of the progressive and the perfect presented above are not directly comparable, since the formulae behind these rates have different denominators (the number of required contexts of the progressive or the perfect in the case of undergeneralisation and all other remaining finite verb phrases in the case of overgeneralisation); therefore, a comparison between the ratios of overgeneralised and undergeneralised progressive and perfect verb phrases in proportion to the total number of progressive and perfect verb phrases in the subcorpora *BUCLE*_110,000 and *GICLE*_110,000 is illustrated in figure 8.4:
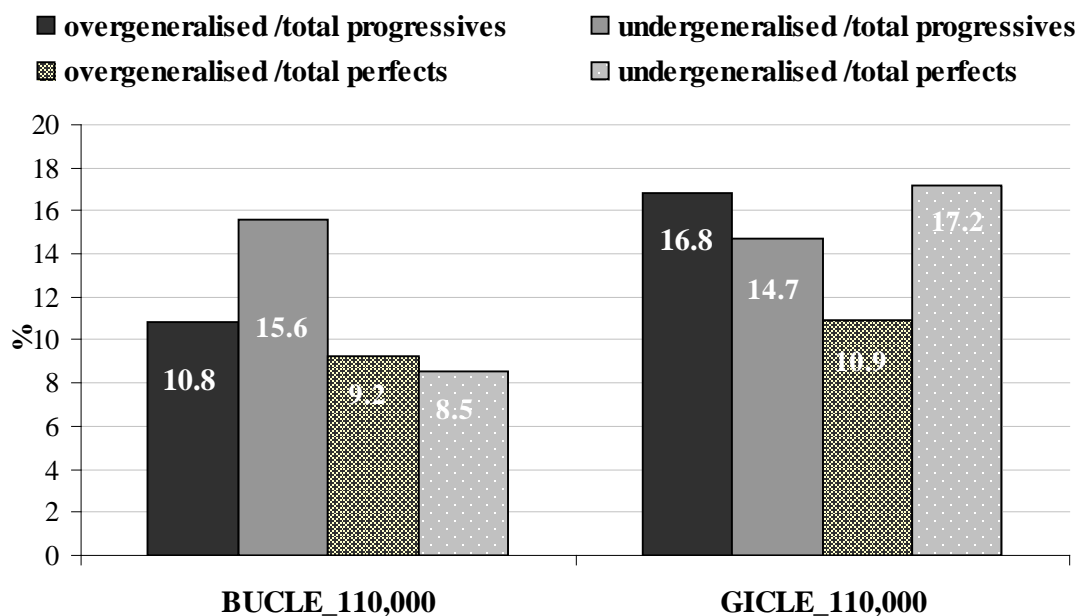
Figure 8.4. Over- and undergeneralised aspect forms in *BUCLE*_110,000 and *GICLE*_110,000 in proportion to the aspect forms in total

One difference concerning the proportion of undergeneralised progressive forms in *BUCLE*_110,000 and *GICLE*_110,000 becomes obvious in figure 8.4: whereas the formula measuring undergeneralised forms in relation to the number of required contexts for these forms presented earlier yielded slightly higher results for *GICLE*_110,000 in terms of the percentage of undergeneralised progressives, the above formula shows that, contrary to these former results, Bulgarian EFL learners undergeneralise progressive forms to a somewhat greater extent than German EFL learners. Nevertheless, no statistically significant differences between *BUCLE*_110,000 and *GICLE*_110,000 in terms of the non-use of the progressive in required progressive contexts were established by this last measurement either.

To summarise, the quantitative measurements of advanced EFL learners' misuse presented in this section can serve as a good way of comparison between the rates of misuse in different learner corpora in addition to the frequency measurements like overuse and underuse traditionally employed within the CIA framework (cf. Granger 1996; Gilquin 2008). In quantitative terms, the analysis in this section revealed that Bulgarian EFL learners show a quantitative preference for the perfect, whereas German EFL learners show a quantitative preference for the progressive, thus confirming the quantitative results on the basis of the whole corpora *BUCLE* and *GICLE* presented in chapter 6; in addition, the comparison between Bulgarian and German EFL learners' rates of overgeneralisation of the progressive

and the perfect and their non-use in required contexts revealed further differences between the two learner populations with regard to the strong undergeneralisation or non-use of the perfect, and especially the present perfect in required present perfect contexts in the German learner corpus, as well as the high rates of undergeneralised progressives in the Bulgarian learner subcorpus. In terms of overgeneralisation of the two aspect forms, the progressive turned out to be extended to non-progressive contexts more often by German EFL learners than by Bulgarian EFL learners, whereas no considerable differences between the two learner groups were found in the case of the overgeneralisation of the perfect and its non-targetlike use in e.g. simple past contexts. These differences, together with the common difficulties for both learner groups will be dealt with in more detail in the next section.

One last point concerning the quantitative measurement of the overall targetlike use of the progressive and the perfect needs to be addressed here: a quantitative comparison between the learner and native corpora on the basis of the normalised frequencies (per 1,000 words) of the targetlike uses of the progressive and the perfect in the learner corpora (i.e. the overall frequency of the progressive and the perfect minus the instances of overgeneralisation in *BUCLE*_110,000 and *GICLE*_110,000) yields lower, yet almost identical results to the frequency results presented in chapter 6 (figure 8.5).
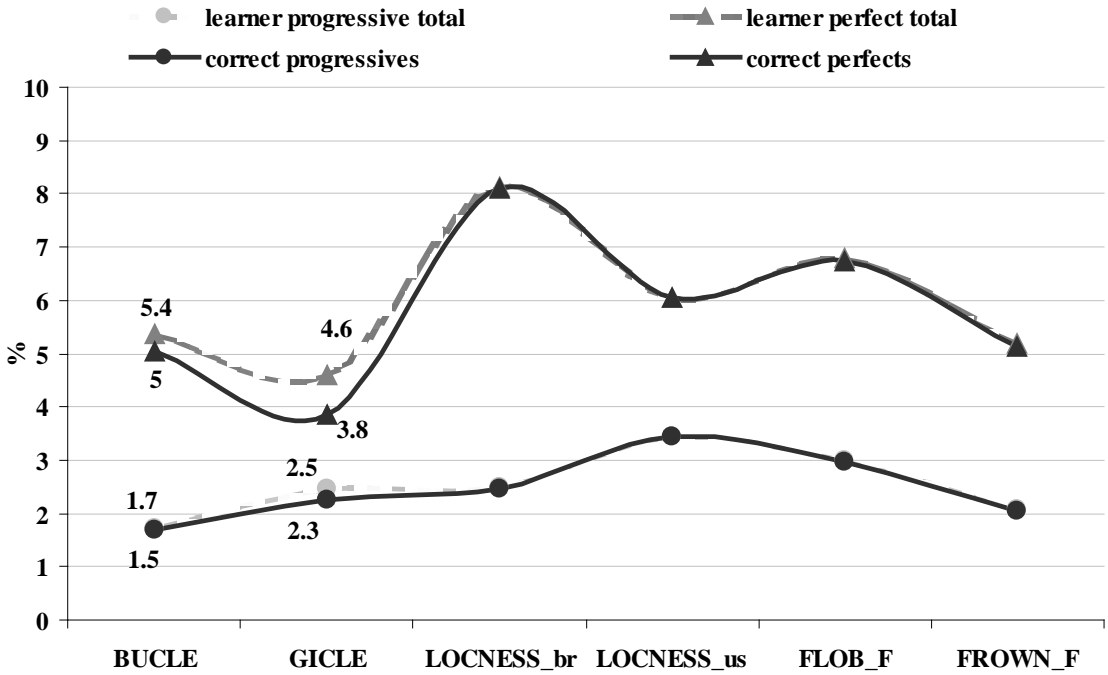
Figure 8.5. Comparison between the normalised frequencies of the progressive and the perfect in total and the normalised frequencies of the targetlike progressives and perfects

Thus, considering the targetlike use of the progressive and the perfect only, it turns out that, on average, Bulgarian and German EFL learners underuse the progressive and the perfect even to a greater extent than the novice native writers from *LOCNESS*_br, *LOCNESS*_us, as well as the expert native writers from *FLOB*_F and *FROWN*_F; moreover, since only half of each of the learner corpora have been error-tagged for errors on the verb phrases, further instances of learner misuse in the remaining, non-error-tagged parts of *BUCLE* and *GICLE* are more than likely, possibly resulting in an even greater underuse of these two aspect forms. The reasons behind this underuse, as well as behind the general patterns of misuse of the progressive and the perfect in advanced Bulgarian and German EFL writing will be contrastively discussed in the final chapter of the present study.

8.6. Summary

With the problem-oriented Computer-Assisted Error Analysis (CEA) of advanced Bulgarian and German EFL learners' misuse of progressive and perfect verb forms, the present chapter offered an alternative approach to the contrastive interlanguage quantification methods used for the identification of non-targetlike aspect use in learners' argumentative writing presented in chapters 6 and 7. The findings supplement the results of the CIA analysis in the sense that they provide a deeper insight into "what learners get right as well as what they get wrong" (Ellis and Barkhuizen 2005: 70). The combined methods of CEA and TLU of learner data yielded both predicted and surprising results with respect to the hypothesised misuse types for the two learner populations in the present sample: first, contrary to the fourth claim of the Aspect Hypothesis (and as already discussed in chapter 7), both Bulgarian and German EFL learners overextend the progressive incorrectly to non-progressive contexts, using it with stative verbs and resulting states, as well as in generic and habitual situations. Similarly, both learner groups overgeneralise the perfect to "classic" simple past contexts, using it "freely" alongside past-time adverbials such as "ago" and "yesterday".

Still, much more interesting than the overgeneralisation of progressive and perfect verb forms to non-progressive and non-perfect contexts are the instances of undergeneralisation or sheer avoidance of both verb forms by Bulgarian and German EFL learners, which the present methodology was able to uncover. Such instances merit a closer investigation, since they possibly signal an incomplete acquisition of the progressive and the

perfect or at least an incomplete form-function mapping in terms of aligning the appropriate aspect forms with the correct aspect values. Here, there are some notable differences between the two learner populations: to illustrate, contrary to the expectations that German EFL learners would "fall back" on the principles of use of the German *Perfekt* as a general narrative tense and be misled to use the equivalent English present perfect form to narrate past events (see also chapter 2), German EFL learners rather avoid using the English present perfect altogether – to a much greater extent than Bulgarian EFL learners. In contrast, Bulgarian EFL learners avoid using the progressive in required contexts or use it very sparingly, replacing it by e.g. the simple present and thus reducing the already small number of progressive verb forms in *BUCLE* even more. Last, but not least, taking into account only the number of target-like progressive and perfect verb forms in the two learner corpora reveals, once again, that both Bulgarian and German EFL learners underuse the two aspect forms highly significantly in comparison to all four native corpora used as a reference in the present study. These findings will be discussed in greater detail in chapter 9.

9. Discussion

The present study builds upon previous research on EFL learners' use of tense-aspect forms in the sense that it combines a frequency-based learner corpus approach (cf. Gilquin 2008, Lenko-Szymanska 2007, Granger 1996; 1998; 1999; Axelsson and Hahn 2001, Virtanen 1997 etc., see also chapters 4 and 6) with a more qualitative, SLA-research-oriented analysis of advanced EFL learners' writing, along the lines of recent SLA studies on learner tense-aspect use like Meunier and Littre (2013), Davydova (2011), Eriksson (2008) and Housen (2002a; 2002b) (see also chapters 3, 4, 7 and 8). The frequency-based analysis of aspect use by advanced Bulgarian and German EFL learners serves as "the starting point of cross-corpus comparison[s]" (Hewings and Hewings 2005: 84) and of more detailed analyses of the distribution patterns of learners' aspect forms across Vendler's (1957) categories of lexical verb types, across clauses and in combination with temporal adverbs and contracted auxiliary forms in learner and native novice and expert writing (see chapter 7). The quantifying method used is a novel adaptation of Smitterberg's (2005) V-coefficient measurement which normalises the frequency of progressive and perfect forms in relation to the number of finite verb phrases (that in theory could have been marked for the perfect and/or the progressive) in essay writing of advanced Bulgarian and German EFL learners in comparison to that of English and American student and expert writers. This quantifying procedure thus yields more fine-grained results which complement the normalisation measures (e.g. per 1,000 words) used by previous learner-corpus studies and employed as reference by the present study too (see chapter 6, sections 6.2. and 6.3.). In addition, the methodology applied in the qualitative part of the study offers an innovative take on traditional Error Analysis, since it combines a problem-oriented (cf. McEnery et al. 2006: 43), computer-assisted error analysis (cf. Dagneaux et al. 1998; Dagneaux et al. 2008) and a subsequent target-like-use and scoring analysis of learners' aspect forms (cf. Pica 1983: 70 – 71). This combined method allows for an in-depth analysis of "what learners get right as well as what they get wrong" (Ellis and Barkhuizen 2005: 70).

On the basis of two subcorpora (*BUCLE*_110,000 and *GICLE*_110,000) tagged for inappropriate tense-aspect uses by a native informant – American English speaker (unaware of the precise research question), the latter method helped to uncover instances of

overgeneralisation of the progressive and the perfect to non-progressive and non-perfect contexts, as well as, more importantly, instances of non-use of the progressive and the perfect in required progressive or perfect contexts in advanced EFL learners' writing. This method proved a valuable addition to the quantitative, distributional analysis offered by chapters 6 and 7 and delivered unpredicted results which help to shed light on the nature of advanced learners' misuse of the progressive and the perfect.

In the following, I will discuss the results for the variation patterns of progressive and perfect verb phrases in advanced Bulgarian and German EFL learners' writing in the light of learner-related variables like the L1 influence, the exposure to English as a target language, as well as learning-related variables like the transfer of training and the L2 writing proficiency. The final section of this chapter will propose a unified model of analysis of aspect use in advanced EFL learners' writing.

9.1. Aspect Variation Patterns in Learner Writing: Some Remarks

The quantitative results for the use of the progressive and perfect forms in advanced Bulgarian and German EFL learners' written English are surprising in the sense that the two learner groups deviate from the corpus-based target norm in a somewhat different manner than proposed by the findings of earlier learner corpus studies (see section 4.3). In a nutshell, although *GICLE* learners use significantly more progressives than *BUCLE* learners, they do not overuse them (as previously reported) in comparison to the majority of the native-speaker English control corpora used as benchmarks in the present study; rather, *GICLE* learners underuse the progressive significantly. The adapted V-coefficient measurement developed specifically for this study to compare the ratios of progressive verb phrases per finite verb forms across the learner and native corpora also corroborates this finding. The results for the perfect reveal a converse, but parallel picture: although *BUCLE* learners use significantly more perfects than *GICLE* learners, they still underuse them highly significantly in comparison to all four native corpora, so that both learner groups, and in particular *GICLE* learners, avoid using the perfect, wherever they can. Nonetheless, learners' sparing use of both aspect forms in quantitative terms is already symptomatic not only of non-targetlike use, but also of an incomplete acquisition of the functions and meanings of the progressive and the perfect, and their subsequent inappropriate use, i.e. of an incomplete "form-to-function mapping" (Housen 2002a: 156). The latter finding implies that the form-to-function mapping

of English aspect forms is incomplete even at an advanced proficiency level such as C1 of the CEFR (accounting for an estimated majority of both learner groups), and even more strikingly, at the mastery or most proficient C2 level (accounting for an estimated 10% of the Bulgarian learners in *BUCLE* and an estimated 35% of the German learners in *GICLE*, or almost twice as high as the *ICLE* average, cf. Gilquin and Granger 2011; Granger et al. 2009).

Incidentally, another look at the distribution of finite verb phrases and in particular simple present verb phrases (not at the focus of the present investigation) shows that both Bulgarian and German EFL learners overuse simple present verb phrases in comparison to all four novice and expert native corpora highly significantly (the majority at the threshold of 0.01% level), with a ratio of about 70% simple present verb phrases of all finite verb phrases in *BUCLE* (15,272 simple present verb phrases out of 22,412 finie verb phrases in total) and 60% simple present verb phrases (14,303 simple present verb phrases out of 24,610 finite verb phrases in total) of all finite verb phrases in *GICLE*. In contrast, the native corpus with the higherst proportion of simple present verb phrases *LOCNESS_us* (9,310 out of 16,474 finite verb phrases in total) features much less simple present verb phrases 56.5% (figure 9.1).
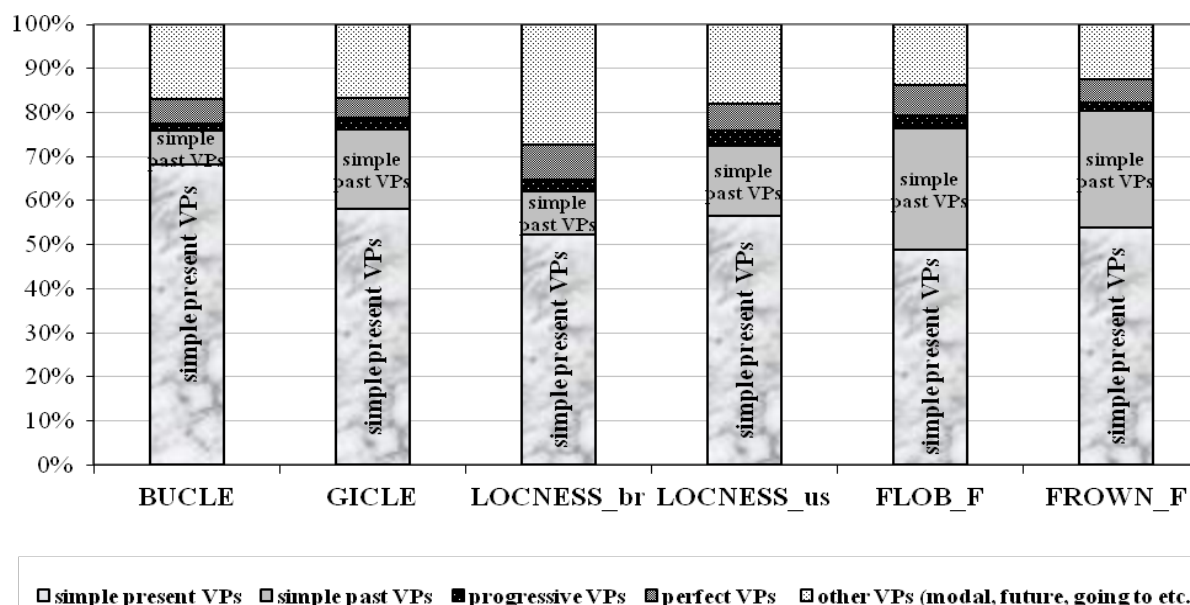
Figure 9.1. Distribution of simple present and simple past verb phrases across the learner and native corpora (see also chapter 6, figure 6.5)

The overuse of simple present verb phrases in the learner corpora has two possible implications:

(1) Bulgarian and German EFL learners use the simple present as the safest, most prototypical form, often incorrectly replacing other, more complex tense-aspect forms by it

(2) There are significant genre differences between the learner and the native corpora which also reinforce a greater variation of verb forms

The first argument is further corroborated by the fact that, as indicated above, an incomplete form-function mapping is manifested in the case of progressive and perfect verb phrases which are much more rarely used in quantitative terms in the learner corpora than in the native corpora. *BUCLE* and *GICLE* learners' avoidance of more complex tense-aspect forms like the present progressive or the present perfect is thus likely to be compensated by learners' overwhelming overuse of the simple present: by frequently using the unmarked, often bare simple present verb form, learners avoid taking greater risks with tense-aspect markings they feel unsure of in terms of form, meaning or use. In other words, learner texts "give the appearance of [using] deliberate present tense continuity" (Godfrey 1980: 108). This linguistic behaviour not only signals an incomplete form-function mapping with respect to tense and aspect in general, but it also fails to meet the expectations for advanced (C1 and C2 level) learners of English who are supposed to be in or beyond the last stage of the acquisition of tense-aspect forms and functions (cf. Housen 2002a, see also section 3.3.2).

The second argument has already been touched upon in sections 5.5 and 6.1: although all six corpora in the present study have been selected for their expository, non-technical style (e.g. only argumentative essay prompts in the case of the learner and novice native corpora, popular commentaries on "current affairs" from the 1990s in the case of *FLOB*_F and *FROWN*_F), no ideal corpus comparability can be guaranteed, since e.g. the number and nature of essay prompts differ significantly across the learner and novice native corpora, as well as the topics of the different commentaries in the expert native corpora. A detailed genre analysis in accordance with Biber's (1988) multidimensional approach which proves the degrees of genre variability in the six corpora goes beyond the scope of the present study; however, we can safely assume that *BUCLE* has the smallest genre variability with only 4 different essay prompts (see section 5.5); and unsurprisingly, it also features the highest percentage (almost 70% of all finite verb phrases) of simple present verb phrases, significantly more than all other corpora including *GICLE* (p<0.01). *GICLE* features fewer simple present verb phrases at the expense of e.g. a higher number of simple past verb phrases

(e.g. induced by personal stories and narrations); furthermore, the distribution of simple present and simple past verb phrases differs across the native corpora too. Although the degree of genre variation across the six corpora used in the present study remains unresolved, it is not necessarily the sole explanation for the greater variability of verb forms in the native corpora. A further argument lending support to the smaller genre variability in the learner corpora are the results for the type-token ratios of progressive and perfect verb phrases across the six corpora: the expert native corpora *FLOB*_F and *FROWN*_F feature the highest type-token ratios or the most varied lexical verbs in the progressive and the perfect, while the learner corpora *BUCLE* and *GICLE* the lowest (see section 7.2). Higher variability on the lexical level may not necessarily go hand in hand with greater variability of the use of grammatical markers; however, it can be assumed that learners' limited vocabulary in the case of the "most prototypical" verbs in the progressive and the perfect (see also Virtanen 1997; Axelsson and Hahn 2001) alongside their overwhelming use of the "most prototypical" tense-aspect form (i.e. the simple present) not only signal a non-targetlike performance, but possibly also an incomplete acquisition of the fully-fledged verb system in English. Both explanations confirm previous research claims that certain verb features remain "unmastered" even at an advanced level – in particular aspect, followed by tense (cf. Meunier and Littre 2013: 68). Thus, the results for the Bulgarian and German EFL learners of the present study (who have been ranked as C1 and partially even as C2-level learners) only corroborate Meunier and Littre's (2013) findings.

In terms of the Aspect and Discourse Hypotheses with respect to the distribution of progressive and perfect verb phrases in learner and native writing, three major trends have been revealed by the present study:

(1) both learner groups and in particular *GICLE* learners **show a stronger preference for atelic** and especially activity **verbs in the progressive** than the native writers of the present study;

(2) both learner groups **use "incorrectly overextended" progressives** with stative verbs, thus contradicting the fourth claim of the Aspect Hypothesis that "[p]rogressive markings are not incorrectly overextended to stative verbs [by learners]" (Andersen and Shirai 1996: 533)

(3) both learner groups and in particular *GICLE* learners show a stronger **preference for perfect forms in subordinate clauses** than the native writers of the control corpora,

thereby possibly "using [perfect] verbal morphology to distinguish foreground from background in narratives" (Bardovi-Harlig 1994: 13) in a different manner from native speakers

The first trend echoes the findings of previous SLA studies (e.g. Collins 2002; 2004; Robison 1995; Bardovi-Harlig 2000, Bardovi-Harlig and Reynolds 1995), all of which attest for their learner samples that most of their learners strongly associate the progressive with activity verbs, often "marking lexical aspect redundantly with morphological aspect" (Bardovi-Harlig 2000: 238). Surprising is only the fact that with a share of 57.7% of all progressives reserved for activity verbs (and ca. 10-15% more on average than the native corpus-based norm), *GICLE* learners prefer considerably more activity verbs in the progressive than *BUCLE* learners (see section 7.1). Having in mind that, on average, *GICLE* learners have been ranked as more proficient than *BUCLE* learners (featuring a greater number of C2 learners – see also section 5.4), this stronger distributional bias challenges the argument that the percentage of activity verbs marked for the progressive gradually drops with increasing proficiency (cf. Bardovi-Harlig and Reynolds 1995; Bardovi-Harlig 2000, Housen 2000; 2002 etc.). Thus, other factors (beyond proficiency level) might be at play here and will be discussed in more detail in the next sections.

The second and third tendency merit more detailed investigation: both learner groups "overextend" the progressive to stative verbs, using it in e.g. inappropriate discourse situations and thus contradicting the fourth claim of the Aspect Hypothesis (cf. Andersen and Shirai 1996: 533), a finding which has already been discussed by previous SLA and learner corpus studies (Housen 2002a; 2002b; Eriksson 2008; Hundt and Vogel 2011). With 7 stative verbs out of the top 20 verbs in the progressive (*be, live, sit, think, stand, watch* and *lie*), *GICLE* is leading the way: although verbs like *live* and *stand* are present among the top 20 verbs in the progressive in the native corpora *LOCNESS*_br, *LOCNESS*_us and *FROWN*_F too, *BUCLE* and especially *GICLE* learners tend to combine the progressive with stative verbs to a much greater extent than the native writers of the present sample. While many of these verbs are generally admissible with the progressive and can express temporary states, waxing and waning situations, agentive activities or emotional undertones (cf. Quirk et al. 1985: 206; Biber et al. 1999: 474; Huddleston and Pullim 2005: 167), a number of the progressive examples with stative verbs from *BUCLE* and *GICLE* were error-tagged as incorrectly

overextended by the native informant (see section 8.3). These examples include overextension of the progressive to stative verbs belonging to three major categories:

(1) verbs of cognition (e.g. *think, perceive* as in e]verybody [...] is probably **thinking** my father is crazy; *<ICLE*-GE-AUG-0016.3> or "when the child **is already perceiving** the things around him" *<ICLE*-BG-SUN-0069.1>);

(2) relational verbs (e.g. *have, depend* as in "they **are depending** on the money from the tourism" *<ICLE*-GE-AUG-0081.1> and "those people who **are actually having** some relation to their relations" *<ICLE*-GE-AUG-0033.3>);

(3) locative verbs in non-temporary, non-dynamic situations (e.g. *sit, lie* as in "animosity and hatred **are lying** in man itself." *<ICLE*-GE-DRE-0024.1>)

The incorrect overextension of the progressive to stative verbs echoes Eriksson's (2008: 192) learner corpus study results for the Swedish sample of *ICLE* with respect to Swedish EFL learners' difficulties differentiating between the uses of e.g. *live* in its progressive form and *live* in its simple form and learners' infelicitous combinations of the progressive "with state verbs [when] the dynamising features agent activity, waxing and waning situations and temporariness are absent" (ibid.: 2008: 192). While *GICLE* learners in particular make use of the above three categories of incorrect overextension of the progressive to stative verbs, *BUCLE* learners overextend the progressive to stative passive verb phrases (as in "the way study **is being organised"** *<ICLE*-BG-SUN-0071.1> or "everything **is being ascribed** financial value" *<ICLE*-BG-SUN-0233.1>) in an equally non-targetlike manner. Again, the resulting states described by these examples in the passive lack dynamising features like e.g. agentivity and temporariness, so that instead of sounding dynamic and ongoing, Bulgarian learners' use of the progressive here sounds unnatural and implies repetitiveness at best. One possible interpretation of learners' (and *GICLE* learners in particular) inappropriate overextension of the progressive to stative verbs and verb phrases is to treat the progressive as a generally-preferred, universally "overextended" form and as "a kind of continuous aspect without temporal immediacy" (van Rooy 2006: 37), which, at the same time, serves as attention-catching, easily recognised and memorised form which is readily employed by learners because of its "higher communicative value in interaction" (Ranta 2006: 112). This interpretation is in line with previous learner corpus studies focussing on German learners' general overuse of the progressive like Axelsson and Hahn (2001) or Hundt and Vogel (2011) who argue that German learner English is the leading EFL variety

with the most progressives; nevertheless, while Hundt and Vogel (Hundt and Vogel 2011: 158) attest a "stretched tolerance" for the progressive with stative verbs for some ESL varieties in their sample, they reject this stretched tolerance for German learner English and maintain that learners overuse the progressive using the most prototypical and widely-taught verbs.

The results obtained by the present study are slightly different: they indicate that while learners overgeneralise the progressive to a small number of highly frequent activity verbs in generic situations and various non-progressive discourse contexts, they also "stretch" the progressive to stative verb phrases in a likewise non-targetlike manner. At the same time, learners' "universal stretching" of the progressive to non-progressive verbs and contexts cannot compensate for the general non-use of the progressive in comparison to the native corpora, both in terms of frequency and in terms of required contexts – this brings us back to the old argument (e.g. Meunier and Littre's 2013) that aspect remains formally and functionally unmastered even at a very advanced proficiency level.

The third tendency concerns Bulgarian and especially German EFL learners' preference for perfect verb phrases in subordinate clauses – while there are virtually no differences in the distribution of progressive verb forms across main and subordinate clauses between the learner and native corpora (the progressive is equally used in main clauses by learners and native speakers), the distribution of perfect verb forms varies according to whether writers are EFL learners or native speakers of English. On average, learners use 10% more present and past perfect forms in subordinate *wh-*, *that-*, zero relative and *if-*clauses than the native writers of the present study, mostly in order to provide supporting story material, preceding temporal circumstances, indirect speech or hypothetical conditions. Given that the majority of these subordinate clauses render the background of the story (cf. Couper-Kuhlen 1994: 231), it seems that the correlation between the choice of aspect form (simple or perfect) and discourse grounding is stronger in the case of the perfect aspect than in the case of the progressive aspect; moreover, Bulgarian and German EFL learners use the perfect in order to distinguish foreground from background in a different manner from native speakers. Surprisingly, the overwhelming majority of the perfects in *BUCLE* and *GICLE* (e.g. *make, take, happen, become* etc., see also section 7.2) are at the same time highly frequent telic verbs conveying kinetic events (cf. Hopper 1989: 216) which are normally used to move a narrative forward, but which – in our case – are used by learners in almost 50% of all cases in

subordinate, backgrounded clauses – the fact that learners use them in subordinate clauses to a much greater extent than native writers runs counter the expectations of the Discourse Hypothesis that telic verbs receive predominantly perfective markings and are found predominantly in the foreground of a story (e.g. Dry 1983; cf. Hopper 1989). This particular finding indicates that Bulgarian and German EFL learners use the present and past perfect in order to mark the background of a story much more frequently than the native writers of the control corpora; it also justifies Bardovi-Harlig's (1995) interest in subordinate or backgrounded clauses as "promising for the study of perfect and progressive" (Bardovi-Harlig 1995: 285). One possible explanation for learners' (and especially *GICLE* learners') preference for the present and past perfect in subordinate clauses is yet again to be found in the genre differences between the learner and native corpora: although produced in response to largely argumentative essay prompts, both *BUCLE* and *GICLE* feature narrations or descriptions which often employ the simple past (in *GICLE* in particular, see also figure 9.1) or the simple present (in *BUCLE* in particular, see figure 9.1) to render the foreground of the story, thereby reserving the present and past perfect for relative and subordinate clauses in order to relate supporting material, even though the uses of the present and past perfect may not be targetlike in these contexts (see also section 8.4.1). In short, while both learners and native speakers foreground the progressive to an almost equal extent (see also Couper-Kuhlen 1994), they differ with respect to the discourse functions of the perfect, which is more often backgrounded by both learner groups, and in particular German EFL learners.

Notwithstanding these results, a word of caution is in order here: both the across-category analysis of lexical verb types in the progressive and the perfect and the analysis of the distribution of progressive and perfect forms across main and subordinate clauses have their limitations: the across-category analysis accounts only for the progressive and perfect verb forms which are states, activities, accomplishments or achievements, disregarding thus the distribution of Vendler's aspectual categories beyond these two morphological markers. A complete analysis of all finite verb phrases in the learner and native corpora may reveal different, i.e. more significant results with respect to the distribution of tense-aspect markers across Vendler's lexical verb types – it may well be the case that e.g. activity verbs in the learner corpora receive much more often progressive markings than activity verbs in the native corpora (for a full comparison between the across-category and within-category analysis see also Bardovi-Harlig 2000; 2002). Likewise, a comprehensive analysis of the distribution of progressive and perfect verb forms alongside all other finite verb forms across

main and subordinate clauses in the native and learner corpora would possibly yield different results with respect to learners' and native speakers' preference for tense-aspect forms to mark discourse grounding. Unfortunately, such a comprehensive analysis goes beyond the scope of the present study which focuses on the progressive and the perfect in particular. The next section will deal with the crucial research question in the present study – the influence of the native language and its measurable effects on the use of aspect forms in advanced Bulgarian and German EFL learners' written English.

9.2. L1 Influence on the Use of Aspect in Bulgarian and German EFL learners' English: Tracing the Untraceable?

One of the most difficult tasks in all SLA research – especially at an advanced stage of learning – is to trace back the origins of the native language L1 in the realm of L2 production, irrespective of developmental patterns or other learning and learner-related factors. This section will attempt to summarise and discuss the findings of the present study in the light of Bulgarian and German as L1s and their influence on the use of advanced EFL learners' use of English aspect morphology in writing. Responding to Shirai's (2009) call for interlanguge comparisons which focus on "the differences between [e.g.] […] German learners (no progressive) vs. [other learners] in the acquisition of the highly grammaticized, polysemous English progressive […]" (Shirai 2009: 184), the present study compares German EFL learners' interlanguage (no grammaticalised progressive, formally fairly similar, but functionally different perfect) with Bulgarian EFL learners' interlanguage (little formal similarities, systemic grammatical opposition between imperfective and perfective verb stems, a grammatical opposition between past aorist vs. past imperfect, perfect vs. evidential category).

A quick glance at the quantitative results for learners' use of the progressive (chapters 6 and 8) reveals that although both learner groups underuse it in comparison to the native control corpora used in the present study, they still behave in a different manner compared with each other – while German EFL learners clearly show a preference for the progressive, overgeneralising it more often to non-progressive contexts and verbs in comparison with Bulgarian EFL learners, Bulgarian EFL learners rather avoid it, also in required contexts and even in combination with adverbials concurrent with the moment of speaking like *now* or

nowadays. Although both native languages lack the progressive as a distinct grammatical category (see chapter 2), it seems thus that the absence of a grammaticalised progressive in L1 German encourages German EFL learners to use the English progressive in expository writing, whereas the presence of a systematic grammatical opposition between perfective and imperfective verb stems alongside an opposition between past aorist and past imperfect in L1 Bulgarian confuses and even discourages Bulgarian EFL learners to use the English progressive at all.

Looking at the lexical verb types after Vendler (1957) in the cases of avoidance of the present progressive and its replacement by the simple present in *BUCLE*_110,000 reveals that the overwhelming majority of these verbs are accomplishment and achievement verbs (e.g. *become, come, develop, start* etc. see examples 8.25 to 8.28 in section 8.3.2) – the assumption that Bulgarian learners would strongly rely on the inherent ongoingness and dynamicity of activity verbs (mostly) to realise progressivity at the expense of progressive markers cannot be fully confirmed here. This finding suggests that the influence from L1 Bulgarian seems to determine the extremely sparing use of the progressive in *BUCLE*: Bulgarian EFL learners fail to provide the necessary progressive markers and rely on the simple present form instead because they seem to equate the respective English simple present verb forms with the equivalent Bulgarian imperfective verb stems which are the unmarked variants in L1 Bulgarian. Indeed, a further analysis of the cases of avoidance of the progressive in the error-tagged subcorpus *BUCLE_110,000* following Gilquin's (2008) and Granger's (2002) Integrated Contrastive Model reveals that in 100% of the cases where the present progressive would have been the native choice, but was incorrectly replaced with the simple present by Bulgarian EFL learners, the corresponding translation equivalent in L1 Bulgarian would have been the present imperfective verb stem[64] (figure 9.2).

---

[64] The translation equivalents of the examples of progressive non-use in L1 Bulgarian were carried out by the author of the present study who is a native speaker of Bulgarian
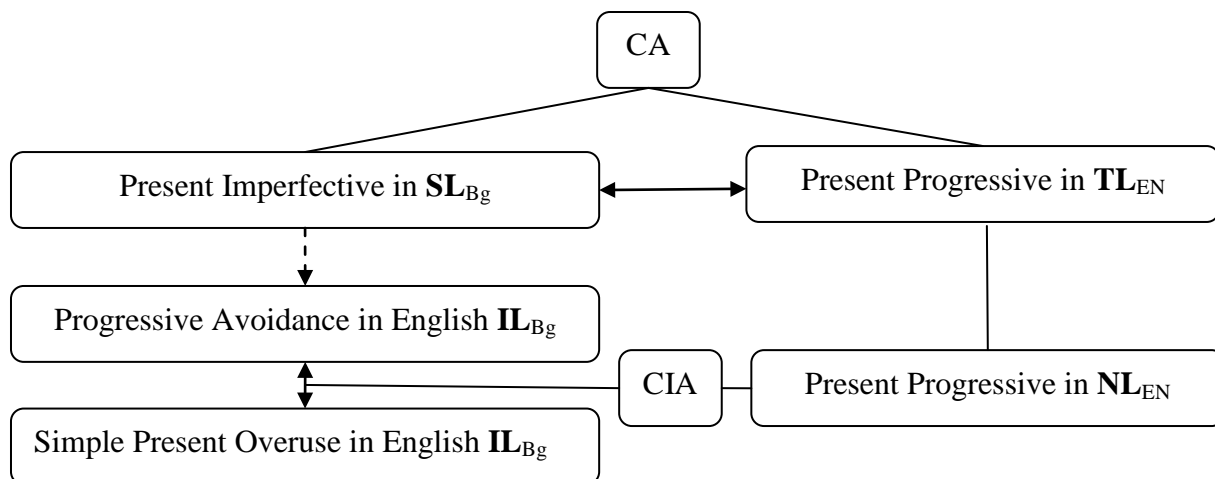
Figure 9.2. Progressive avoidance in *BUCLE*_110,000 and corresponding translation equivalents in L1 Bulgarian

The overwhelming overuse of simple present verb forms in *BUCLE* mentioned in the previous section corroborates this finding; thus, it seems plausible that Bulgarian EFL learners subconsciously feel that there is no "need" for the progressive marker in those cases, since they seem to equate the meaning of English verbs with imperfective verb stems in L1 Bulgarian, which are much more "deeply rooted in the lexicon" (Bertinetto and Delfitto 2000: 190), and distinctly different from the English periphrastically-realised progressive aspect.

Contrary to expectations, German EFL learners – whose native language lacks the progressive as an institutionalised grammatical category, alongside further markers of imperfectivity like in L1 Bulgarian – use the progressive much more frequently than Bulgarian EFL learners, mostly with atelic verbs (activities and states), and also incorrectly overextend it to e.g. activity verbs in non-dynamic, non-progressive contexts and to stative verbs in inadmissible combinations. The latter finding supports Collins and Izquierdo's (2008) and Housen's (2002a; 2002b) claim that the absence of a grammatical aspect may increase learners' reliance on the inherent temporality in the meaning of verbs – since German EFL learners' L1 lacks the progressive, they demonstrate a stronger reliance on the inherent temporality (e.g. ongoingness and dynamicity) of activities (but also states) to realise progressivity. Analysing the equivalent German translations of the avoided progressives in *GICLE* in accordance with Gilquin's (2008) and Granger's (2002) Integrated Contrastive Model would, however, be pointless, since the unmarked equivalent German form in written German will in any case be the *Präsens* form and not one of the emergent periphrastic constructions realising progressivity which have not yet been institutionalised (cf. König and Gast 2009, see also section 2.3). In contrast to *BUCLE*, where all undergeneralised

progressives are present progressives, *GICLE* also features several cases of undergeneralised past progressives, which have been replaced by the simple past (and in only three cases by the simple present) by German EFL learners. At the same time, although German EFL learners use the progressive much more easily and frequently than Bulgarian EFL learners, they still underuse it in comparison to the native corpora used as benchmarks for the present study, which can hardly be explained only in the light of transfer from L1 German.

With regard to Bulgarian and German EFL learners' use of the perfect, the reverse picture becomes obvious: although both learner groups underuse perfect forms significantly in comparison to almost all native corpora (with the exception of *FROWN*_F in the case of *BUCLE*), there are again clear differences between Bulgarian and German EFL learners: Bulgarian EFL learners use considerably more perfect forms than German EFL learners, also overgeneralising them (and in particular the present perfect) to non-perfect contexts to a greater extent than *GICLE* learners. In sharp contrast, the ratio of undergeneralised perfects in *GICLE*_110,000 is almost double the ratio of undergeneralised perfects in *BUCLE*_110,000 (see sections 8.4.2 and 8.5) – a finding which clearly contradicts all expectations that the formal similarity between the German *Perfekt* and the English present perfect would serve as the basis for German learners' frequent replacing of the simple past with the present perfect, as learners would have done in e.g. narratives in their native L1 German (cf. Comrie 1976; König and Gast 2009; Löbner 2002; Klein 2000; Klein and Vater 1998 etc.). Rather, German learners avoid using the present perfect altogether, thus confirming Davydova's claim that German EFL learners "[b]eing unsure of the exact meanings conveyed by the English perfect […] try to avoid using this form altogether, replacing it with a semantically simpler form – the preterite" (Davydova 2011: 288).

One striking difference between the two learner subcorpora *BUCLE*_110,000 and *GICLE*_110,000 concerns the avoidance of the present perfect and its replacement by other verb forms: whereas in the German learner subcorpus the majority of the cases of avoidance of the present perfect feature its replacement by the simple past (almost 78%), the Bulgarian learner subcorpus exhibits greater variation and much fewer instances of replacement of the present perfect by the simple past (52.7%, see figure 9.3).
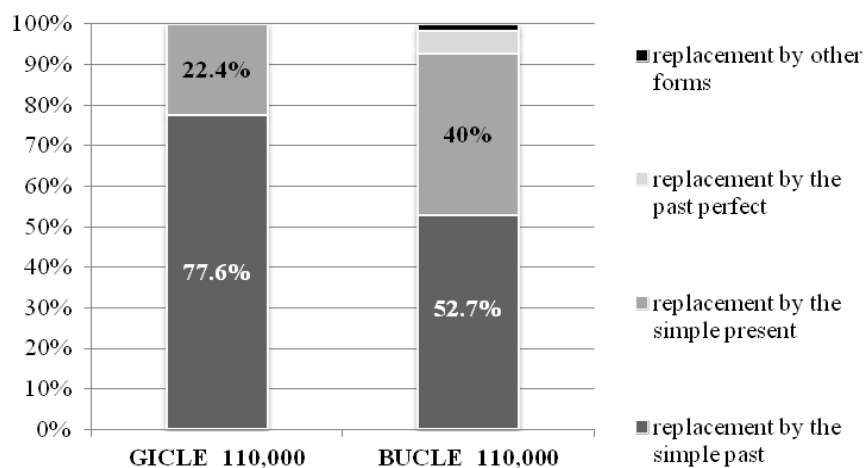
Figure 9.3. Present perfect avoidance and replacement by other verb forms in *GICLE*_110,000 and *BUCLE*_110,000

This particular finding suggests that the association between the meanings of the present perfect with past temporality seems to be much stronger for *GICLE* learners than for *BUCLE* learners; in contrast, *BUCLE* learners associate the meanings of the present perfect with both the simple past and the simple present, thus replacing it in written English alternatingly by the two forms. Incidentally, a closer look at the instances of avoidance of the present perfect and its replacemenet by other verb forms in the two learner subcorpora and a consequent analysis of their translational equivalents in L1 Bulgarian and L1 German[65] in accordance with Gilquin's (2008) and Granger's (2002) Integrated Contrastive Model confirms the assumption that *GICLE* learners associate the present perfect more strongly with past temporality than *BUCLE* learners: while 43.9% of the translational equivalents to the avoided present perfects in *BUCLE_110,000* involve the use of the present (imperfective and perfective) stem, only 15.8% of the translational equivalents of the avoided present perfects in *GICLE*_110,000 would be in *Präsens* in L1 German (see figure 9.4).

---

[65] The German translation equivalents of the examples of present perfect non-use in *GICLE*_110,000 were carried out by three informants – native speakers of German, who were given the choice between two translation variants (one featuring the *Präteritum* and one the *Perfekt*) and asked to select the most natural-sounding one; the translational equivalents of the present perfect avoidance examples in *BUCLE*_110,000 were carried out by the author
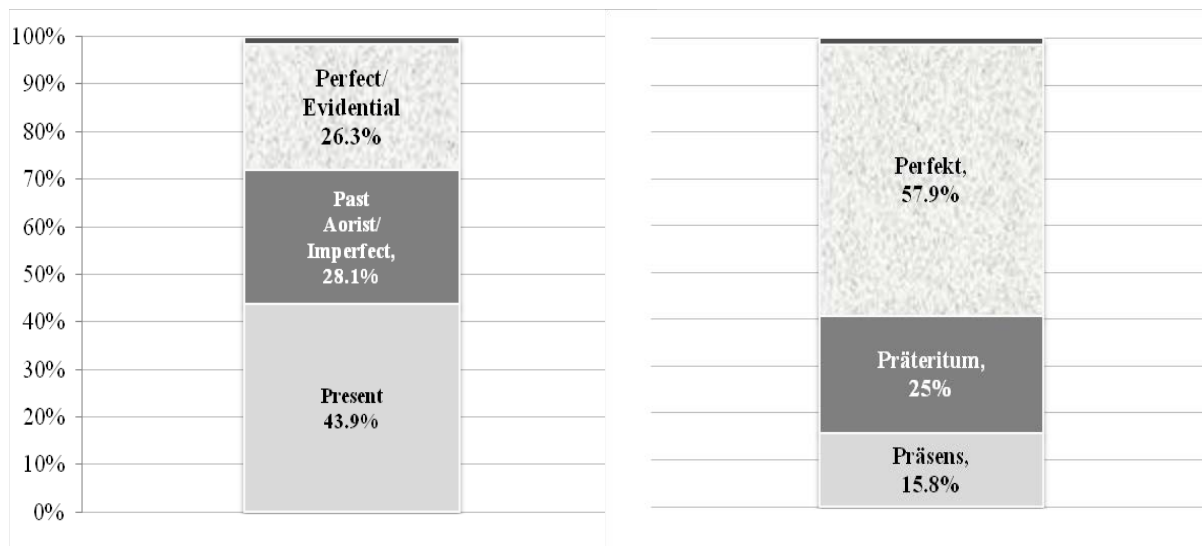
Figure 9.4. Present perfect avoidance in *GICLE*_110,000 and *BUCLE*_110,000 and corresponding translation equivalents in L1 German and L1 Bulgarian

Interestingly enough, in about 58% of the cases where *GICLE* learners replaced the present perfect by other forms (e.g. the simple past), they would have most probably used *Perfekt* in their native L1 German, rather than *Präteritum*, as indicated by the native German informants; at the same time, most cases in which the native German informants opted for translational equivalents involving the *Präteritum* feature either verbs in the passive (both *Vorgangspassiv* and *Zustandspassiv*) or the verbs be (*sein*) and have (*haben*) in L1 German, which are more commonly used in *Präteritum* and not in *Perfekt* in standard written German (e.g. König and Gast 2009). In short, German EFL learners fail to differentiate between the meanings of the morphologically-marked forms simple past and present perfect in English, since the parallel opposition *Präteritum - Perfekt* in their native L1 German does not make a great difference to them with regard to "the temporally encoded meaning [in L1 German]" (Hahn 2007: 57) – especially considering the increasing use of the German *Perfekt* form as a non-perfect, narrative past (cf. Löbner 2002: 388; Stechow 2002: 393). At the same time, contrary to expectations, rather than overgeneralising the present perfect to simple past contexts in English to a much greater extent, German EFL learners avoid using the present perfect and replace it by the simple past in classic present perfect contexts involving anteriority and recent result, even if modified by adverbial phrases stereotypically used with the present perfect (see section 8.4.2). In contrast, Bulgarian EFL learners use more present perfect forms than German EFL learners in total and avoid them to a lesser extent; at the same time, they seem to confuse the English present perfect much more frequently with the simple present than German EFL learners – in particular in current relevance contexts, where they

234

would have predominantly used a present form (both perfective and imperfective, and often passive) in their L1 Bulgarian (see also section 8.3.2). In other words, the native language seems to play a role with regard to learners' temporal conceptualisation or the meanings learners associate with the present perfect – while German EFL learners' confusion of the present perfect with the simple past reflects to a certain extent their temporal conceptualisation in L1 German, Bulgarian learners' perception and interpretation of the English present perfect is more ambiguous: Bulgarian EFL learners associate (and confuse) the present perfect with both present and past temporality. The differences between the two learner groups in terms of avoidance of the present perfect signal possible L1-induced difficulties in the acquisition of the core meanings of the perfect: while German EFL learners have troubles differentiating between recent or indefinite past and definite past due to L1-induced levelling of the meaning opposition, Bulgarian learners seem to confuse the meanings of the present perfect in a broader sense, i.e. they fail to use the perfect in both indefinite past and recent past with current relevance contexts, possibly due to L1-related conceptualisations of the English perfect as a form with a wide variety of meanings similar to the Bulgarian perfect's "bifurcated ...[developmental path as both] experiential and non-narrative [...] and indirective and narrative" (Lindstedt 2000: 377).

This being said, both learner groups overgeneralise the present perfect predominantly to simple past contexts, especially in narratives and in descriptions framed by the simple past, *BUCLE* learners doing so to an even greater extent than *GICLE* learners (see sections 8.4.1 and 8.5). Here, L1-transfer resulting from the formal similarity in L1 German seems likely: *GICLE* learners use the present perfect freely in narratives instead of the simple past, also in combinations with "classic" adverbs signalling definite past moments such as "ago" and "yesterday". However, much more interesting is the fact that *BUCLE* learners overuse the present perfect in simple past contexts in an even more uninhibited manner than *GICLE* learners: to illustrate, over 90% of the overgeneralised present perfects in *BUCLE*_110,000 are in simple past contexts; moreover, a quick glance at the translational equivalents of the incorrectly overgeneralised present perfects in *BUCLE*_110,000 in L1 Bulgarian shows that 100% of them involve use of the Bulgarian perfect, both in its indefinite past and evidential meanings (figure 9.5).
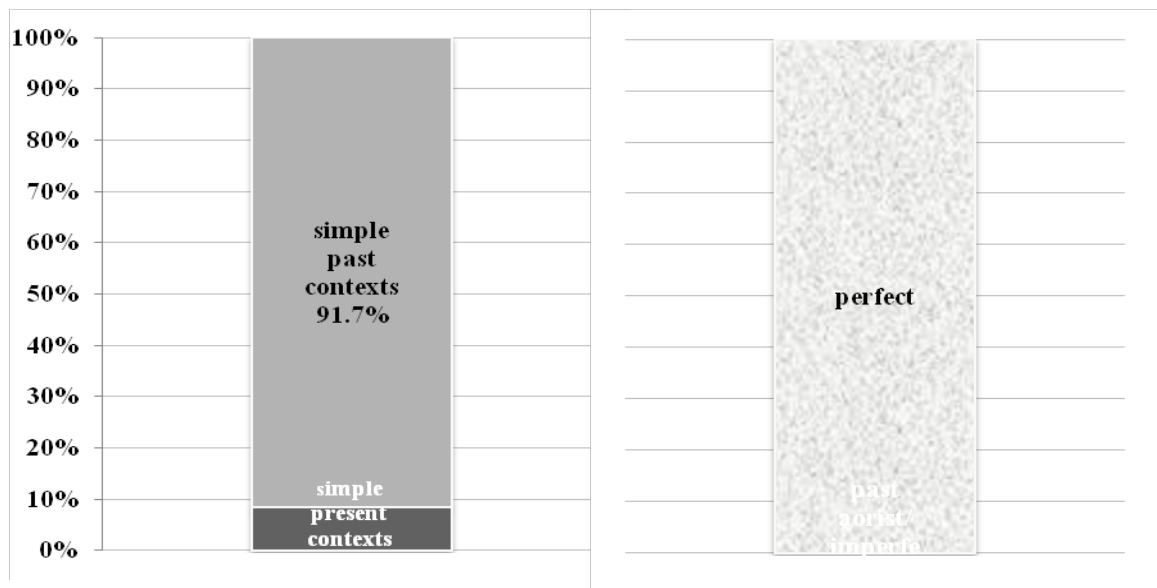
Figure 9.5. Present perfect overgeneralisation in *BUCLE*_110,000 (left) and corresponding translation equivalent forms in L1 Bulgarian (right)

The latter finding clearly signals conceptual transfer from L1 Bulgarian: in addition to Bulgarian EFL learners' uncertainty of the exact range of meanings of the English present perfect, there comes also a misinterpretation of the function of the present perfect as a narrative tense into play: BUCLE learners transfer the functions of the Bulgarian perfect as an indirect category of reported forms typical of "second-hand" narratives, scientific articles and reports whose authors quote the findings of other scholars (cf. Fici 2005: 39) to English in an attempt to signal lack of "first-hand" evidence about facts in their argumentation. A quick glance at the overused present perfects in BUCLE_110,000 proves that the majority of the overused present perfect forms involve third person subjects (singular and plural), which suggests that the Bulgarian essay authors relate stories which they have not experienced first-hand and which, to them, require the use of the present perfect as a form of "hearsay". In other words, the majority of incorrectly overused present perfects in BUCLE function as non-targetlike modality markers conceptually transferred from L1 Bulgarian, rather than as aspectual markers.

To conclude, the presence or absence of a formal similarity between e.g. the German *Perfekt* and the English present perfect does not translate straightforwardly into negative transfer which affects the results observably and significantly (cf. Davydova 2011) and accounts for the differences between the two learner groups as a sole factor; rather, Bulgarian and German EFL learners often overgeneralise the "rule[s] and principles of the target language without reference to the L1 system, a process that [is] called 'internal interference'"

(Nemser 1971 in Sharwood Smith 1994: 37). Still, Bulgarian EFL learners tend to have more troubles with the acquisition and targetlike use of the progressive, while German EFL learners have more troubles with the targetlike use of the perfect. In the case of Bulgarian EFL learners' limited use and sheer avoidance of the progressive, L1 interference can be established when comparing the instances of replacemenet of the progressive in required contexts with their translational equivalents (present imperfective forms) in L1 Bulgarian; in the case of German EFL learners' avoidance of the progressive, the lack of a progressive category in L1 German influences only an increased reliance on the inherent durativity in the meaning of verbs – both states and activities. Such comparisons prove efficient also in the case of Bulgarian EFL learners' misuse of the perfect, which is avoided in contexts where learners have failed to acquire finer meaning nuances idiosyncratic for English, but overused in contexts in which learners transfer their L1 conceptualisation of the perfect as "the required" verb form in narratives and arguments not evidenced by themselves. In contrast, the expected interference from the parallel German *Perfekt* form resulting in a hypothesised overwhelming substitution of the English simple past by the present perfect is much less significant than expected and confirms Davydova's (2011) findings with respect to German EFL learners' insecurity about the present perfect and its persistent avoidance; nevertheless, this avoidance indicates fear to use the present perfect which can at least be partially explained with transfer-induced fear (see the next section).

Notwithstanding these findings, it is important to note that "[t]ransfer is a slippery phenomenon that does not lend itself easily to apprehension" (Gilquin 2008: 25) and is often untraceable, especially in advanced EFL learners' output; therefore, tracing the influence of the native language requires contrastive evidence from many different datasets along the lines of Gilquin's (2008) and Granger's (2002) Integrated Contrastive Model. Admittidely, there is no one-to-one relationship between e.g. the replaced aspect forms in the learner corpora and the corresponding translational equivalents in L1 German or L1 Bulgarian; in addition, both error-tagging and translation are subjective and cannot reconstruct what learners actually "wanted" to say. Therefore, the final section of this chapter will attempt to present a unified model of analysis of aspect use in advanced EFL learners' writing. The next section will illuminate some further aspects influencing learner use of the progressive and the perfect beyond transfer from the native language.

9.3. Further Factors: L2 Proficiency, Writing Expertise, Exposure and Transfer of Training

The present section examines the findings with respect to further learner-related variables influencing advanced Bulgarian and German EFL learners' use of aspect beyond transfer from the native language like proficiency and L2 writing expertise, L2 exposure and possible teaching-induced effects. Chapter 6 revealed that both learner groups and the native student writers in the present sample move along a formality-colloquiality continuum which reflects their proficiency and writing expertise in expository writing and which is at the same time related to the orality-literacy continuum proposed by Koch and Oesterreicher (1985). Moreover, *BUCLE* was found not only to be the "verbiest" corpus of all six corpora with the highest ratio of finite verb phrases (see section 6.1), but also the corpus with the most simple present verb phrases (see section 9.1), the least progressive verb forms and the lowest TTR ratio for lexical verb types in the perfect. Considering the fact that *BUCLE* learners were rated as less proficient (e.g. less C2 learners) than *GICLE* learners by independent raters, we can safely assume that Bulgarian EFL learners' general non-targetlike use of the progressive and the perfect has a lot to do with their lower proficiency and reflects an incomplete acquisition of the fully-fledged verb system in English. Even though somewhat more advanced than *BUCLE* learners, *GICLE* learners equally stand out with their "verbiness", low TTR for lexical verb types in the progressive, significant avoidance of the perfect and general non-targetlike use of the progressive (e.g. extension to stative verb phrases) and especially of the perfect.

In addition to learner proficiency, there are clear differences with regard to EFL learners' and native writers' writing competence: we can observe a similar cline when comparing the ratios for the finite verb phrases across the six corpora (from high to low), the distribution of simple present verb phrases (from high to low), the TTRs for the progressive and the perfect (from low to high), as well as the TTRs for adverbial phrases modifying the progressive and the perfect (from low to high): the two learner corpora *BUCLE* and *GICLE* and the two expert native corpora FLOB_F and FROWN_F form almost invariably the two opposite ends of the scale, with the novice native corpora usually in-between.

```
┌─────────────┐                                              ┌──────────────────┐
│  Orality    │                                              │   Scripturality  │
└─────────────┘                                              └──────────────────┘
```

Learner Corpora > Novice Native Corpora > Expert Native Corpora

L2 Writing Expertise
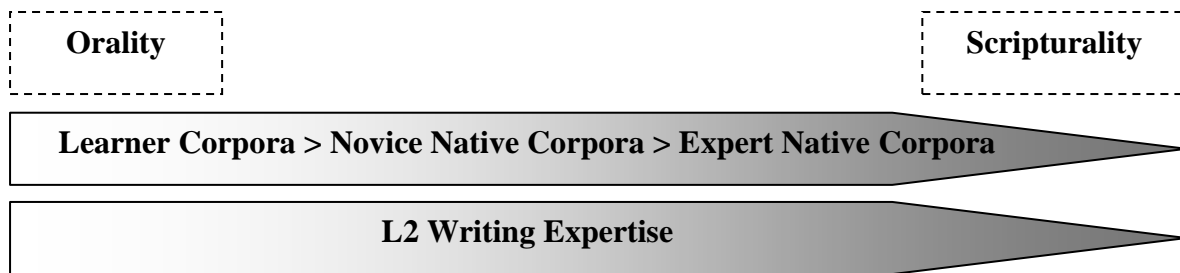
Figure 9.6. Distribution of the learner and native corpora along a writing expertise cline

To a certain extent, this cline can be explained with increasing writing expertise: it reflects Koch and Oesterreicher's (1985) orality-literacy continuum and ties in with Eriksson's (2008), Altenberg's (1997) and Gilquin and Paquot's (2007) findings that advanced learners of English use a number of colloquial features in their written production which are more typical of speech than of academic writing and which also signal lack of L2 writing competence, apart from contributing to a greater colloquial overtone of learner texts in comparison to (especially expert) native-speaker texts. Interestingly enough, when looking at the contracted auxiliary forms used with the progressive and the perfect as typical oral features, we observe a breach of the orality-literacy continuum proposed above: even though the expert writers of *FROWN*_F are among the most experienced in writing, they use the most auxiliary contractions, followed by the *GICLE* and *BUCLE* learners, *LOCNESS*_us, *FLOB*_F and lastly, the inexperienced writers of *LOCNESS*_br.

This finding is striking as it shows a certain regularity in the way learner language patterns with either British or American English – both British corpora *LOCNESS*_br and *FLOB*_F feature lower contraction rates than the American corpora *LOCNESS*_us and *FROWN*_F; *BUCLE*'s rates are lower and closer to the British corpora, whereas *GICLE*'s rates are higher and closer to the American corpora. The higher use of auxiliary contractions in written English reflects a "tendency for spoken language habits to infiltrate the written language" (Leech and Smith 2006: 198) and especially American English usage which is "leading the way" towards colloquialisation (cf. Leech and Smith 2006: 199). The total lack of occurrences of contracted auxiliaries in the British novice native corpus *LOCNESS*_br reveals either a stronger affinity of British novice writers to the prescriptive norm, or a possible insecurity or lack of expertise in using more varied language, since *FLOB*_F writers, who are experienced journalists, use more auxiliary contractions than *LOCNESS*_br writers. Likewise, *GICLE* learners, who have been rated as more advanced learners, use more

contracted auxiliaries: therefore, a multitude of factors such as learner proficiency, writing expertise and the type of native variety are at play here.

With respect to the latter factor, the present study observed an interesting trend concerning learners' orientation towards either British or American English: there are clear similarities between German learner English and American English and Bulgarian learner English and British English, both in terms of the use of the progressive and the use of the perfect: in terms of their relative frequencies, German learners and American native speakers prefer the progressive over the perfect, whereas Bulgarian learners and British native speakers prefer the perfect over the progressive. Likewise, the learner corpora pattern with either native corpus-based norm in terms of orality features like contracted auxiliaries: German learners and American novice and expert writers use more auxiliary contractions with the perfect and the progressive, whereas Bulgarian learners and British novice and expert writers use much less or none contracted auxiliaries with both the perfect and the progressive. In other words, since the vast majority of *GICLE* learners (see figure 9.7) have had greater exposure to English in an English-speaking environment and with native English teachers in the classroom (cf. Granger et al. 2002; 2009; Lorenz 2002: 102 in Granger et al. 2002), they also feel much more comfortable using e.g. colloquial features in writing.
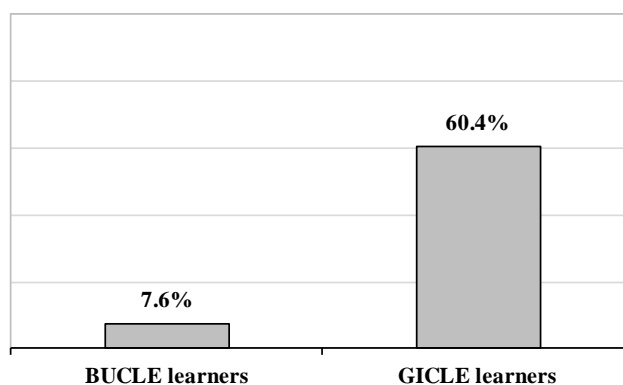
Figure 9.7. Percentage of learners who stayed at least 1 month abroad in an English-speaking country

At the same time, the patterning of German learner English with American English, rather than British English indicates that the L2 exposure for German EFL learners may have possibly been American English, rather than British English; unfortunately, *ICLE* reveals no information about the English-speaking countries students stayed in. In contrast, very few

Bulgarian EFL learners (under 10%) in the 1990s[66] – when the Bulgarian part of *ICLE* was sampled – had the chance to stay in an English-speaking country – i.e. *BUCLE* learners had much less exposure to any kind of native English, apart from the "poverty-of-stimulus-ridden" Bulgarian EFL classroom. Therefore, the similarity between e.g. Bulgarian EFL learners' and British native writers' frequencies for the progressive and the perfect may be due to Bulgarian EFL learners' stronger orientation to the prescriptive norm as the only type of exposure provided in an "impoverished" EFL environment in the 1990s – which was British English (cf. Blagoeva 2002).

Last, but not least, some of the instances of inappropriate use of the progressive and the perfect in *BUCLE* and *GICLE* indicate possible transfer-of-training effects (cf. Selinker 1972; 1976; Gass and Selinker 1994), i.e. either mirrored or misinterpreted features of the input created by "teacher[s] or textbook[s]" (Sharwood Smith 1994: 37). For instance, the persistent avoidance of the present perfect by German EFL learners also signals that there may have been an overemphasis on the present perfect as a complex and a risky form in the German EFL classroom due to the formal similarity between the German *Perfekt* and the English present perfect, resulting thus in German learners' raised awareness and transfer-induced fear to use the English present perfect altogether (e.g. Davydova 2011). In addition, both learner groups' inability to maintain tense continuity, often demonstrated by random replacement of the simple past and the simple present by other tense-aspect forms (see chapter 8) or unmotivated tense shifts between e.g. the past perfect and the simple past (section 8.4.1) may well echo the inadequate, sentence-based, rather than discourse-based representations of tense and aspect in English textbooks (cf. Eriksson 2008; Granger 1999). Likewise, learners' preference for temporal modification of the progressive (especially Bulgarian EFL learners) and the perfect (especially German EFL learners) suggests a possible teaching-induced bias due to an overemphasis of the most prototypical adverbs accompanying the progressive and the perfect in the foreign language classroom and ELT materials (cf. Davydova 2011; Römer 2005; Mindt 2000) leading to an overuse of temporal specification that misleads learners to believe that once they use the "typical" adverbs, the tense-aspect forms accompanying them are bound to be targetlike[67].

---

[66] Both Great Britain and the USA required visa and proof of funds for students applying for an exchange term in the 1990s, which was a major obstacle for most Bulgarian students to go abroad

[67] The present chapter has focused on learners' proficiency, writing expertise and transfer of training, since a detailed investigation of the influence of ELT materials and the representations of the progressive and the perfect in ELT textbooks along Römer's (2005) research lines would have proven difficult with respect to the lack of information on the textbooks used by ICLE learners; therefore such an investigation has not been undertaken

The final section of this chapter will attempt to present an integrated model of analysing aspect use in L2 argumentative writing alongside a summary of the results.

9.4. Towards a Model of Analysing Aspect Use in EFL Writing

The previous three sections discussed the findings on aspect variation patterns in advanced Bulgarian and German EFL learners' writing from the perspective of a multitude of factors ranging from learner-related variables like the L1 influence and the L2 proficiency to learning-related variables and developmental patterns. The present section offers a brief synthesis of the results which aims at proposing a model for the analysis of aspect use in L2 argumentative writing integrating the methodological frameworks employed by the present study alongside a schematic representation of the most significant results for advanced Bulgarian and German EFL writing on the basis of this model.

In a recent book on research design and methodology to tense and aspect, Salaberry et al. (2013) criticise that so far, "research in tense and aspect has used both qualitative and quantitative methods, but very few researchers have combined them" (Salaberry et al. 2013: 439) and call for a mixed methodology approach to L2 tense-aspect research. The present study is an attempt at such a "mixed-method experimentalism" (Salaberry et al. 2013: 439) since it employs a multi-level analysis of learner use of aspect in English which includes a quantitative phase establishing contrasts between learner and native frequencies (e.g. Granger 2009), followed by a qualitative phase which focuses on specific details such as lexical co-selection and discourse distributions of aspect forms which help explain the observed quantitative tendencies (cf. Dörnyei 2007: 45). The qualitative part combines an annotation of lexical verb types in the progressive and the perfect and a computer-assisted "problem-oriented annotation" (McEnery et al. 2006: 43) involving error tagging of verbs in two learner subcorpora carried out by a native informant. A subsequent quantitative retrieval of the error tags and a targetlike-use analysis after Pica (1983) closes the loop back to the initial frequency analysis by providing a comparison between the relative learner frequencies of aspect forms in total and the relative frequencies of targetlike uses of the progressive and the perfect. I call this multi-layer analysis an integrated model for the analysis of aspect use – which is diagrammatically represented in figure 9.8:
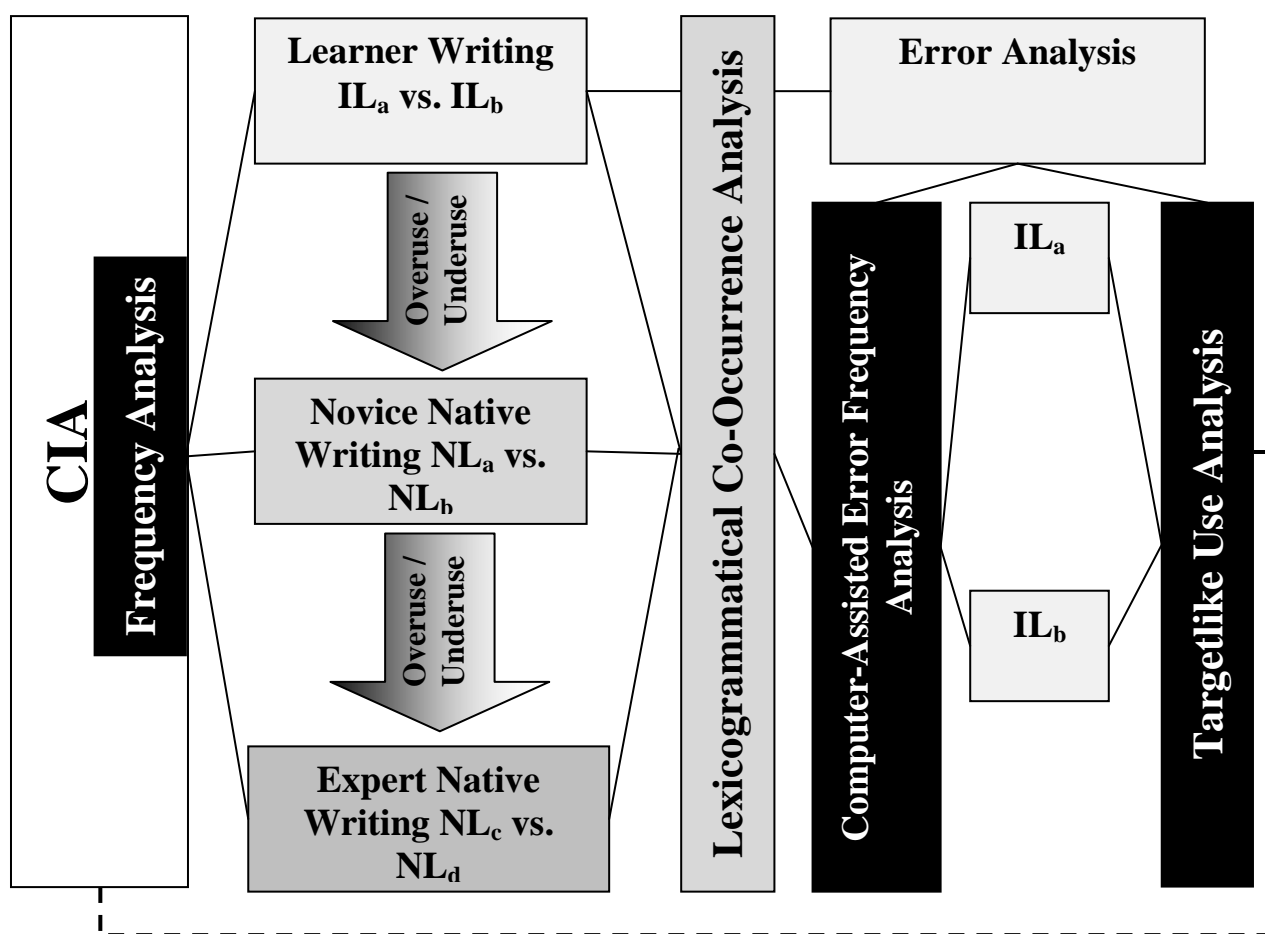
Figure 9.8. An integrated model for the analysis of aspect use in EFL writing

The proposed model for analysing learner aspect considers: a) the frequency-based analysis (involving the three datasets consisting of learner writing IL_a and IL_b, novice native writing NL_a vs. NL_b and expert native writing NL_c and NL_d; b) the lexicogrammatical analysis of the co-occurrence of grammatical aspect forms with lexical aspect verb types, c) a Computer-Assisted Error Analysis (CEA) involving error frequency counts and d) a Targetlike-Use Analysis (TLU), which takes into account both instances of overgeneralisation and non-use of aspect forms and which helps explain the quantitative tendencies revealed by the initial frequency analysis. The integrated model is innovative in the sense that it responds to Salaberry et al.'s (2013) concerns as it attempts to combine both quantitative and qualitative perspectives to learner use of aspect in EFL writing and thus to explain and enrich the insights gained through the "classic" CIA used as the basis for the present model. In addition to the quantifying methods used by most studies adopting the CIA approach, the present learner corpus study employs a novel quantifying method – an adaptation of Smitterberg's (2005) V-coefficient measurement which normalises the frequency of

progressive and perfect forms in relation to the number of finite verb phrases (that could in theory have carried wither perfect and/or progressive marking). This quantifying procedure yields more fine-grained results which complement the normalisation measures utilised by previous studies; moreover, the advantage of extracting and analysing finite verb phrases in POS-tagged corpora is that it allows for cross-linguistic comparisons of e.g. finite verb trends in different interlanguages in comparison with different varieties of the same target language. Combined with a frequency analysis of the progressive and perfect, the ratios of contracted auxiliaries used with the progressive and the perfect, the TTRs for progressive and perfect verb phrases etc. across the six corpora (see previous section), the present method mirrors Eriksson's (2008), Altenberg's (1997) and Gilquin and Paquot's (2007) results and can be utilised to place the results on a two-dimensional scale following both Koch and Oesterreicher's (1985) orality-scripturality continuum and Biber et al.'s (1999) quantitative trends for American English and British English – as presented in figure 9.9:
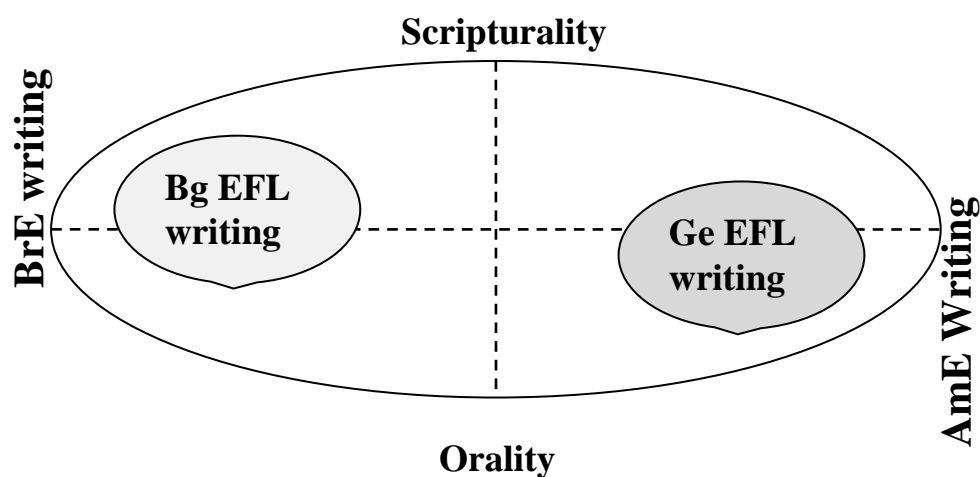
Figure 9.9. Two-dimensional scale of Bulgarian and German EFL writing

The two-dimensional scale summarises the results of the present learner corpus study with respect to orality-scripturality features in learner and British and American novice/expert native writing, as well as the quantitative trends with respect to the preference for the progressive and the perfect in Bulgarian and German interlanguage in relation to American and British English (cf. Biber et al. 1999: 461 – 465); i.e., German EFL writing is somewhat more colloquial or "oral" in nature than Bulgarian EFL writing, and also much closer to American English frequencies for both the perfect and the progressive. In contrast, Bulgarian EFL writing is slightly less colloquial (more limited use of e.g. auxiliary contractions among others) and deliberate or not, closer to British English corpus-based norms for the progressive

and the perfect. The results for the four native corpora *LOCNESS*_br, *LOCNESS*_us, *FLOB*_F and *FROWN*_F largely confirm previously discussed trends of "colloquialisation" in argumentative writing (cf. Mair and Hundt 1995), which is led by American English usage "leading the way" in grammatical changes in progress (cf. Leech and Smith 2006: 199): the American corpora *LOCNESS*_us and *FROWN*_F are more colloquial (e.g. verbier, using more auxiliary contractions etc.) than the British corpora *LOCNESS*_br and *FLOB*_F.

Finally, the results for Bulgarian and German EFL learners' use of the progressive and the perfect on the basis of the present mixed-method model reveal a moderate "cross-linguistic influence involving relativistic effects" (Odlin 2008: 306), i.e. the influence of the native languages German and Bulgarian is mitigated by other factors such as the classroom environment and the effects of "overteaching", the amount and type of L2 exposure, as well as universal developmental stages such as the influence of lexical aspect. In short, most significant and equally astonishing are the results with respect to the avoidance of the perfect in German EFL writing and the progressive in Bulgarian EFL writing; somewhat less surprising are the results for the overextension of the progressive in German EFL writing and the perfect in Bulgarian EFL writing (see figure 9.10). The L1-L2 formal similarity between the German *Perfekt* and the English present perfect certainly plays a role in terms of the non-targetlike use of the perfect in German EFL writing, although not the expected one: in addition to the instances of overgeneralisation of the present perfect to e.g. simple past contexts in *GICLE* (most probably having an interlingual source), much more significant are the instances of non-use or avoidance of the present perfect – in all likelihood due to learners' teaching-induced fear to use the present perfect in response to teachers' over-emphasis on the perfect as a "risky form" in the German EFL classroom.

Simultaneously, Davydova argues that this particular linguistic behaviour also reflects learners' simplification strategy – i.e. replacing a complex form from the target language with a simpler one (cf. Davydova 2011: 11), which also appears plausible in the case of both Bulgarian and German EFL learners' replacement of the present perfect by the simple past. Further, most problematic for Bulgarian EFL learners is the use of the progressive aspect – apart from the simplification strategies mentioned above, Bulgarian learners' persistent avoidance also seems to be due to interlingual transfer from L1 Bulgarian which incorporates progressivity as part of the verb stem – being thus much more integrated in the Bulgarian lexicon than the English periphrastically-formed progressive aspect – and misleading less

advanced Bulgarian EFL learners to believe that progressive markings are "superfluous" since they are "carried" by the verb itself. Conversely, the overgeneralisation of the progressive to non-progressive contexts in German learner English is hardly due to the absent progressive category in L1 German: rather, its overextended use is conditioned by e.g. lexical verb types involving features of durativity or extra-linguistic influences like e.g. the amount of exposure to spoken English, possibly also in an American English-speaking environment. In contrast, the overextension of the perfect to non-perfect contexts in Bulgarian EFL writing has most probably an interlingual source, since the functional differences between the perfect in TL English and the perfect in L1 Bulgarian (e.g. mostly aspectual vs. aspectual as well as modal) lie at the heart of Bulgarian EFL learners' conceptual transfer of the reported, modal functions of the Bulgarian perfect to the English present perfect and Bulgarian learners' consequent non-targetlike use.
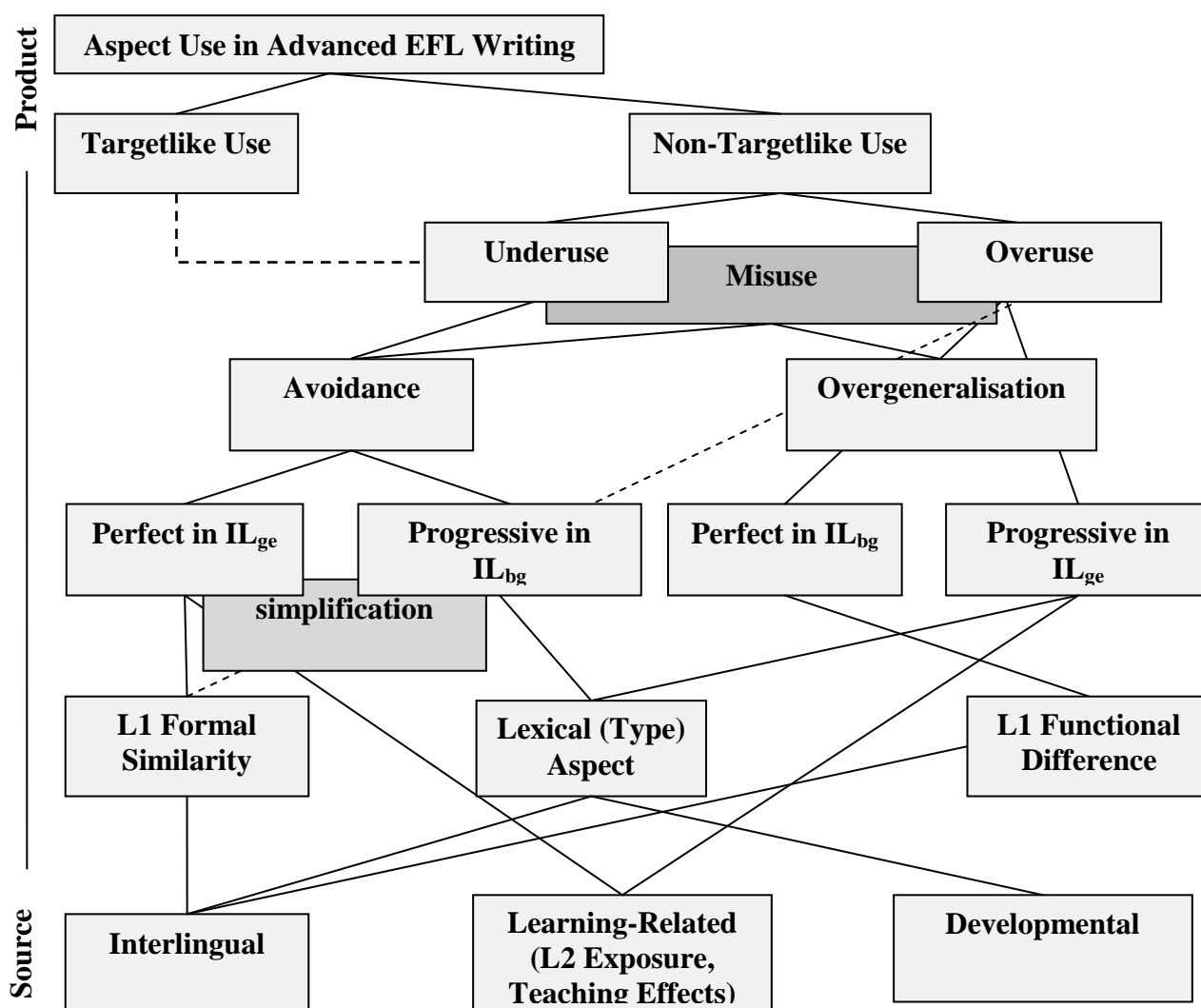
Figure 9.10. Synthesis of the results for aspect use in advanced Bulgarian and German EFL writing

To summarise, the boundaries between targetlike and non-targetlike use are not always clear-cut: to illustrate, learners' targetlike use as revealed by the qualitative analysis employed by the present study constitutes at the same time significant underuse in quantitative terms; moreover, the target hypothesis for one learner expression is bound to vary, since it remains, in all likelihood, just one out of many. The last chapter of the present study will round-up the discussion with a brief conclusion and suggestions for future research.

10. Conclusion and Prospects for Future Research

The immediate objective of the present study has been to account for similarities and differences in the use of English aspect forms in the writing of two learner populations with different aspectual systems in their native languages that have not been compared thus far and remained a research desideratum (cf. Shirai 2009: 184) – Bulgarian and German EFL learners of English. To this end, both learner corpus data from the Bulgarian and German parts of *ICLE* and native-speaker corpus data from *FLOB*, *FROWN* and the Louvain corpus of novice native writing *LOCNESS* were drawn for a Contrastive Interlanguage Analysis of the progressive and the perfect in expository writing. The results for the quantitative distribution of aspect forms in learner writing show significant deviance from the native-speaker corpus-based norm: both learner populations underuse the progressive and the perfect, although to a different extent: Bulgarian EFL learners underuse the progressive (and especially the present progressive) more significantly, whereas German EFL learners rather underuse the perfect (and in particular the present perfect). Taking into consideration the novel quantification method employed in the present study (i.e. the adapted V-coefficient ratio of the number of progressive and perfect forms to the total number of finite verb phrases), the study uncovered that learners' underuse of both aspect forms is even more pervasive when comparing the frequencies of finite verb phrases across the learner and native corpora. It seems reasonable to suggest that both avoidance/simplification strategies and the L1 influence are at play here: in the case of Bulgarian EFL learners' underuse of the present progressive, both conceptual transfer based on the L1-induced misinterpretation of progressivity as part of the lexicon and simplification strategies due to a lower proficiency level of Bulgarian EFL learners are plausible; in the case of German EFL learners' underuse of the present perfect, the L1-induced insecurity about when to use a form similar to the German *Perfekt*, but different in terms of meaning, is also possibly reinforced by an over-emphasis on the English present perfect in the German EFL classroom as a "risky form". Moreover, simplification seems to be the overarching common strategy for both learner groups to solve these problems: both Bulgarian and German EFL learners overuse the simple present (especially Bulgarian EFL learners) and the simple past (especially German EFL learners) to compensate for the non-use of other, more complex tense-aspect combinations: in this way, learners create a "tense continuity" which is misleading since it conceals persisting difficulties with tense-aspect forms such as the present progressive and the present perfect they have not yet fully mastered.

The quantitative tendencies concerning the distribution of progressive and perfect forms in advanced Bulgarian and German EFL writing were further reinforced by the qualitative, computer-assisted error analysis employed in the second half of the empirical part: such a mixed approach proved valuable for shedding light on the possible reasons behind learners' underuse; in addition, drawing evidence from potential translational equivalents following Granger's (1996) and Gilquin's (2008) Integrated Contrastive Model helped explain or reject assumptions about the native-language influence on learners' non-targetlike use of the progressive and the perfect (e.g. German EFL learners' non-use of the present perfect in required contexts and its replacement by the simple past did not necessarily correspond to L1 preference for the *Präteritum* in the equivalent German translations). Likewise, the targetlike-use quantification measurement proposed in chaper 8 helped to enhance the results by emphasising on the differences between the two learner corpora. In contrast, the lexicogrammatical variation analysis in chapter 7 revealed common difficulties for both learner populations in terms of learners' genre insensitivity and lack of writing competence, learners' preference for durative verbs in the progressive, as well as their preference for the perfect in subordinate clauses and in the "company" of temporal adverbials. Mirroring Eriksson's (2008: 109) findings on advanced Swedish EFL learners' errors, the present study also found a number of middle-ground infelicities in both learner corpora (i.e. non-targetlike uses which are not strictly errors, but which were marked by the native informant as non-nativelike, usually with a comment or an alternative suggestion), especially in the case of overgeneralisation of the progressive to non-progressive contexts. Finally, the proposed orality-scripturality cline offered a good starting point for comparisons between learner data with different sets of native-speaker data – it provided quantitative evidence that writing competence is something to be "learned" which equally holds for both learners of English and inexperienced native British or American writers; in addition, it raised questions about which English variety (e.g. British or American English) learners adhere to more closely (e.g. German EFL learners' aspect use resembling American English use and Bulgarian EFL learners' use resembling British English use)

In sum, the study has raised some interesting points about the corpus-based analysis of learners' non-use which have not yet been looked into; in addition, it has addressed some further methodological issues like the combination between "classic" corpus-based CIA analysis with "classic" SLA techniques like Error Analysis and Targetlike Use Analysis, as

well as the need for taking into account further factors (e.g. the type and amount of L2 exposure) beyond the native language influence used as the basis for all Interlanguage contrasts. Future learner corpus research and especially learner corpus design and compilation should therefore consider tracking down all learner variables in greater detail, and especially focus on variables that may have appeared irrelevant so far (e.g. the English-speaking environment learners spend time in when going abroad, the influence of foreign language teaching, the types of textbooks used in the foreign language classroom etc.). Moreover, I appeal for a more rigid control of variables such as learner proficiency in order to ensure full comparability between individual corpora. Finally, a combination of methods such as a corpus-based contrastive analysis of learners' output and e.g. elicitated samples would yield more fine-grained results.

Naturally, the results of the present study are by no means comprehensive, and only provide a limited description of a small area of English grammar (i.e. aspect) which is inseparable from other areas like tense and modality – future learner corpus research should address e.g. the interplay between tense, aspect and modality in the case of differing L1 systems. Further, a fully error-tagged learner corpus would have been much more valuable for the present analysis; likewise, a within-category analysis of e.g. all activities that have been marked for the progressive, perfect, simple present etc. might have led to slightly different results or intensified the results of the across-category analysis even more.

In short, future learner corpus research should further refine its methods at the interface of SLA research techniques and should also strongly consider experimental data to complement and enhance learner corpus research results (cf. also Meunier and Littre 2013); e.g. by using the same learner sample for elicitation tasks on tense and aspect and for learner corpus compilation. Only a combination of methods can ensure a better understanding of authentic learner language, especially with respect to the reasons and sources behind learners' non-targetlike use in quantitative terms. To summarise, I hope for a bright future of a closer cooperation between learner corpus researchers and SLA researchers working together on the collection, compilation and annotation of learner data in order to provide for a comprehensive and dynamic picture of learner language which takes into account a variety of factors behind second-language use.

References

Abe, M., & Tono, Y. (2005), "Variations in L2 Spoken and Written English: Investigating Patterns of Grammatical Errors across Proficiency Levels" In P. Danielsson & M. Wagenmakers (Eds.), *Proceedings from the Corpus Linguistics Conference Series* (pp. 1–11). Birmingham.

Abraham, W. (1991), "How Much of the German Tense System Is 'Aspect' and 'Aktionsart'?" *Belgian Journal of Linguistics, 6*, 133–150.

Abraham, W., & Janssen, T. (1989), "Einleitung" In W. Abraham & T. Janssen (Eds.), *Linguistische Arbeiten (LArb): 237. Tempus-Aspekt-Modus: Die lexikalischen und grammatischen Formen in den germanischen Sprachen* (pp. 1–23). Tübingen: Niemeyer.

Aijmer, K. (2002), "Modality in Advanced Swedish Learners' Written Interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 55–76). Amsterdam, Netherlands: Benjamins.

Allerton, D. (2008), "Swiss English, German English and American English: In Grammatical Alliance against Traditional British English?" In M. A. Locher (Ed.), *Contributions to the Sociology of Language: Vol. 95. Standards and Norms in the English Language* (pp. 23–45). Berlin etc.: Mouton de Gruyter.

Altenberg, B. (2002), "Using Bilingual Corpus Evidence in Learner Corpus Research" In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 38–54). Amsterdam, Netherlands: Benjamins.

Andersen, R. W. (1984), "The One-To-One Principle of Interlanguage Construction", *Language Learning, 34*(4), 77–95.

Andersen, R. W., & Shirai, Y. (1994), "Discourse Motivations for Some Cognitive Acquisition Principles", *Studies in Second Language Acquisition, 16*(2), 133–156.

Andersen, R. W., & Shirai, Y. (1996), "The Primacy of Aspect in First and Second Language Acquisition: The Pidgin-Creole Connection" In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 527–570). SanDiego etc.: Academic Press.

Anderson, L. B. (1982), "The "Perfect" as a Universal and as a Language-Specific Category", In P. J. Hopper (Ed.), *Typological Studies in Language: Vol. 1. Tense-Aspect: between Semantics & Pragmatics. Containing the Contributions to a Symposium on Tense and Aspect held at UCLA May 1979* (pp. 227–264). Amsterdam etc.: Benjamins.

Andersson, S.-G. (1989), "On the Generalization of Progressive Constructions: "Ich bin (das Buch) am Lesen"-Status and Usage in three Varieties of German" In L.-G. Larsson (Ed.), *Acta Universitatis Upsaliensis: Studia Uralica et Altaica Upsaliensia: 19. Proceedings of the Second Scandinavian Symposium on Aspectology* (pp. 95–105). Uppsala: Almqvist & Wiksell.

Aronson, H. I. (1985), "Form, Function and "The Perfective" in Bulgarian" In M. S. Flier & A. Timberlake (Eds.), *UCLA Slav. Studies: 12. The Scope of Slavic Aspect* (pp. 274–285). Columbus, OH: Slavica.

Axelsson, M., & Hahn, A. (2001), "The Use of the Progressive in Swedish and German Advanced Learner English: a Corpus-Based study", *ICAME journal, 25*, 5–30.

Bailey, N. (1989), "Discourse Conditioned Tense Variation: Teacher Implications" In M. R. Eisenstein (Ed.), *The Dynamic Interlanguage. Empirical Studies in Second Language Variation* (pp. 279–296). New York: Plenum Press.

Bailey, N., Madden, C., & Krashen, S. D. (1974), "Is There a "Natural Sequence" in Adult Second Language Learning?" *Language Learning, 24*(2), 235–243.

Bardovi-Harlig, K. (1992), "The Relationship of Form and Meaning: A Cross-Sectional Study of Tense and Aspect in the Interlanguage of Learners of English as a Second Language", *Applied Psycholinguistics, 13*(03), 253–278.

Bardovi-Harlig, K. (1994), "Anecdote or Evidence? Evaluating Support for Hypotheses Concerning the Development of Tense and Aspect" In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research Methodology in Second-Language Acquisition* (1st ed., pp. 41–61). Hillsdale NJ etc.: Erlbaum.

Bardovi-Harlig, K. (1995), "A Narrative Perspective on the Development of the Tense/Aspect System in Second Language Acquisition", *Studies in Second Language Acquisition (SSLA), 17*(2), 263–291.

Bardovi-Harlig, K. (1995), "The Interaction of Pedagogy and Natural Sequences in the Acquisition of Tense and Aspect" In F. R. Eckman, D. Highland, P. W. Lee, J. Mileham, & R. R. Weber (Eds.), *Second Language Acquisition Theory and Pedagogy* (pp. 151–168). Mahwah, NJ: Erlbaum Ass.

Bardovi-Harlig, K. (1997), "Another Piece of the Puzzle: the Emergence of the Present Perfect", *Language Learning: A Journal of Research in Language Studies, 47*(3), 375–422.

Bardovi-Harlig, K. (1997), "Assessing Grammatical Development in Interactional Contexts", *TESOL Quarterly, 31*(4), 797–806.

Bardovi-Harlig, K. (1999), "From Morpheme Studies to Temporal Semantics: Tense-Aspect Research in SLA", *Studies in Second Language Acquisition, 21*(3), 341–382.

Bardovi-Harlig, K. (2000), *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use. Language Learning.* Malden, Mass.: Blackwell.

Bardovi-Harlig, K. (2002), "Analyzing Aspect" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 129–154). Amsterdam etc.: Benjamins.

Bardovi-Harlig, K. (2005), "Tracking the Elusive Imperfect in Adult L2 Acquisition: Refining the Hunt" In P. Kempchinsky & R. Slabakova (Eds.), *Studies in Natural Language & Linguistic Theory (StNLLT): 62. Aspectual Inquiries* (pp. 397–419). Dordrecht, Netherlands: Springer.

Bardovi-Harlig, K. (2007), "One Functional Approach to Second Language Acquisition: the Concept-Oriented Approach" In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition. An Introduction* (pp. 57–75). Mahwah etc.: Erlbaum.

Bardovi-Harlig, K., & Reynolds, D. W. (1995), "The Role of Lexical Aspect in the Acquisition of Tense and Aspect", *TESOL Quarterly, 29*(1), 107–131.

Baroni, M. (2009), "Distributions in Text" In A. Lüdeling & M. Kytö (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft: Vol. 29. Corpus Linguistics. An International Handbook. Volume 2* (pp. 803–822). Berlin: de Gruyter.

Bertinetto, P. M., & Delfitto, D. (2000), "Aspect vs. Actionality: Why they should be kept apart" In Ö. Dahl (Ed.), *Empirical Approaches to Language Typology (EALT): Eurotyp 20-6. Tense and Aspect in the Languages of Europe* (pp. 189–225). Berlin, Germany: Mouton de Gruyter.

Biber, D. (1987), A Textual Comparison of British and American Writing. *AS, 62*(2), 99–119.

Biber, D. (1988), *Variation across speech and writing.* Cambridge: Cambridge Univ. Press.

Biber, D. (1989), "A Typology of English Texts", *Linguistics, 27*(1 [299]), 3–43.

Biber, D. (1990), "Methodological Issues Regarding Corpus-Based Analyses of Linguistic Variation", *Literary & Linguistic Computing, 5*(4), 257–269.

Biber, D. (2001), "On the Complexity of Discourse Complexity: a Multi-Dimensional Analysis", In S. Conrad & D. Biber (Eds.), *Studies in Language and Linguistics. Variation in English. Multi-Dimensional Studies* (pp. 215–240). Harlow: Longman.

Biber, D. (2006), *Dimensions of Register Variation: A Cross-Linguistic Comparison* (1. pb. version.). Cambridge: Cambridge Univ. Press.

Biber, D. (2006), *University Language: A Corpus-Based Study of Spoken and Written Registers. Studies in Corpus Linguistics: Vol. 23.* Amsterdam: Benjamins.

Biber, D., & Conrad, S. (2001), "Multi-Dimensional Methodology and the Dimensions of Register Variation in English" In S. Conrad & D. Biber (Eds.), *Studies in Language and Linguistics. Variation in English. Multi-Dimensional Studies* (pp. 13–42). Harlow: Longman.

Biber, D., & Conrad, S. (2001), "Quantitative Corpus-Based Research: Much More Than Bean Counting", *TESOL Quarterly, 35*(2), 331–336.

Biber, D., & Conrad, S. (2004), "Corpus-Based Comparisons of Registers" In C. Coffin, A. Hewings, & K. O'Halloran (Eds.), *Applying English grammar. Functional and corpus approaches* (pp. 40–56). London: Arnold.

Biber, D., Conrad, S., & Reppen, R. (1998), *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge etc.: Cambridge Univ. Press.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002), "Speaking and Writing in the University: A Multidimensional Approach", *TESOL Quarterly, 36*(1), 9–48.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999), *Longman Grammar of Spoken and Written English.* Harlow: Longman.

Binnick, R. I. (1991), *Time and the Verb: A Guide to Tense and Aspect.* New York: Oxford Univ. Press.

Brinton, L. J. (1988), *The Development of English Aspectual Systems: Aspectualizers and Post-Verbal Particles. Cambridge Studies in Linguistics: Vol. 49.* Cambridge etc.: Cambridge Univ. Press.

Brinton, L. J. (2000), *The Structure of Modern English: A Linguistic Introduction.* Amsterdam etc.: Benjamins.

Brown, R. (1973), *A First Language: The Early Stages*. Cambridge, Mass.: Harvard Univ. Press.

Burt, M. K., & Kiparsky, C. (1976), "Global and Local Mistakes", In J. H. Schumann & N. Stenson (Eds.), *New Frontiers in Second Language Learning* (pp. 73–79). Rowley, Mass.: Newbury House.

Bybee, J. L. (1985), *Morphology: A Study of the Relation Between Meaning and Form. Typological Studies in Language: Vol. 9.* Amsterdam etc.: Benjamins.

Bybee, J., Perkins, R., & Pagliuca, W. (1994), *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago Ill.: University of Chicago Press.

Collins, L. (2002), "The Roles of L1 Influence and Lexical Aspect in the Acquisition of Temporal Morphology", *Language Learning, 52*(1), 43–94.

Collins, L. (2004), "The Particulars on Universals: A Comparison of the Acquisition of Tense-Aspect Morphology among Japanese- and French-Speaking Learners of English", *Canadian Modern Language Review, 61*(2), 251–274.

Collins, L. (2005), "Accessing Second Language Learners' Understanding of Temporal Morphology", *Language Awareness, 14*(4), 207–220.

Comrie, B. (1976), *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems. Cambridge Textbooks in Linguistics.* Cambridge: Cambridge University Press.

Comrie, B. (1985), *Tense.* Cambridge etc.: Cambridge Univ. Press.

Comrie, B. (2001), "Some Thoughts on the Relation between Aspect and Aktionsart", In A. Barentsen & Y. Poupynin (Eds.), *Functional Grammar: Aspect and Aspectuality, Tense and Temporality. Essays in Honour of Alexander Bondarko* (pp. 43–50). München: Lincom Europa.

Comrie, B. (Ed.) (2002), *The Slavonic Languages* (Repr.). London etc.: Routledge.

Connor, U. (1996), *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing.* Cambridge [u.a.]: Cambridge Univ. Press.

Conrad, S. (2007), "Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century?" In W. Teubert & R. Krishnamurthy (Eds.), *Critical Concepts in Linguistics. Corpus linguistics.* (pp. 55–66). London: Routledge.

Corder, S. P. (1976), "Idiosyncratic Dialects and Error Analysis", In J. H. Schumann & N. Stenson (Eds.), *New Frontiers in Second Language Learning* (pp. 100–113). Rowley, Mass.: Newbury House.

Corder, S. (1991), "The Significance of Learners' Errors", In D. Nehls (Ed.), *Studies in Descriptive Linguistics (SDL): 21. Studies in Contrastive Linguistics and Error Analysis, III: Practical Applications* (pp. 27–35). Heidelberg: Groos.

Couper-Kuhlen, E. (1995), "On the Foregrounded Progressive in American Conversational Narrative: a New Development?" In W. Riehle (Ed.), *Proceedings of the Conference of the German Association of University Teachers of English: Vol. 16. Proceedings of the Anglistentag 1994, Graz* (pp. 229–245). Tübingen: Niemeyer.

Dagneaux, E., Dennes, S., & Granger, S. (1998), "Computer-Aided Error Analysis", *System, 26*, 163–174.

Dahl, Ö. (2000), "The Tense-Aspect Systems of European Languages in a Typological Perspective" In Ö. Dahl (Ed.), *Empirical Approaches to Language Typology (EALT): Eurotyp 20-6. Tense and Aspect in the Languages of Europe* (pp. 3–25). Berlin, Germany: Mouton de Gruyter.

Davydova, J. (2011), *The Present Perfect in Non-Native Englishes: A Corpus-Based Study of Variation. Topics in English Linguistics: Vol. 77.* Berlin etc.: de Gruyter.

Declerck, R. (2006), *The Grammar of the English Tense System: A Comprehensive Analysis. Topics in English Linguistics: Vol. 60,1.* Berlin: Mouton de Gruyter.

Dietrich, R., Klein, W., & Noyau, C. (1995), *The Acquisition of Temporality in a Second Language. Studies in Bilingualism: Vol. 7.* Amsterdam: Benjamins.

Dowty, D. R. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ.* Dordrecht: Reidel.

Dörnyei, Z. (2007), *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies.* Oxford etc.: Oxford Univ. Press.

Dry, H. A. (1983), "The Movement of Narrative Time" *Journal of Literary Semantics, 12*(2), 19–53.

Dulay, H. C., & Burt, M. K. (1974), "Errors and Strategies in Child Second Language Acquisition", *TESOL Quarterly, 8*, 129–136.

Dulay, H. C., & Burt, M. K. (1974), "Natural Sequences in Child Second Language Acquisition", *Language Learning, 24*(1), 37–53.

Ebert, K. H. (2000), "Progressive Markers in Germanic Languages" In Ö. Dahl (Ed.), *Empirical Approaches to Language Typology (EALT): Eurotyp 20-6. Tense and Aspect in the Languages of Europe* (pp. 605–653). Berlin, Germany: Mouton de Gruyter.

Ellis, N. C., & Collins, L. (2009), "Input and Second Language Acquisition: The Roles of Frequency, Form, and Function: Introduction to the Special Issue", *The Modern Language Journal, 93*(3), 329–335.

Ellis, R. (1982), "The Origins of Interlanguage", *Applied Linguistics, 3*(3), 207–223.

Ellis, R. (1987), "Interlanguage Variability in Narrative Discourse: Style Shifting in the Use of the Past Tense", *Studies in Second Language Acquisition, 9*(1), 1–19.

Ellis, R. (1994), *The Study of Second Language Acquisition.* Oxford etc.: Oxford Univ. Press.

Ellis, R., & Barkhuizen, G. (Eds.) (2006), *Analysing Learner Language* (2nd [impression]). Oxford: Oxford University Press.

Elsness, J. (1997), *The Perfect and the Preterite in Contemporary and Earlier English. Topics in English Linguistics: Vol. 21.* Berlin, New York: Mouton de Gruyter.

Elsness, J. (2009), "The Present Perfect and the Preterite" In G. Rohdenburg & J. Schlüter (Eds.), *Studies in English language. One Language, two Grammars? Differences between British and American English* (1st ed., pp. 228–246). Cambridge: Cambridge Univ. Press.

Eriksson, A. (2008), *Tense and Aspect in Advanced Swedish Learners' Written English. Acta Universitatis Gothoburgensis: Vol. 92.* Göteborg: Göteborgs Univ.

Erling, E. J. (2002), "'I Learn English since Ten Years': The Global English Debate and the German University Classroom", *English Today: The International Review of the English Language, 18*(2 [70]), 8–13.

Fici, F. (2005), "The Grammatical Strategies of the Bulgarian Perfect and Its Modal Meanings", *Supostavitelno Ezikoznanie/Sopostavitel'noe Jazykoznanie/Contrastive Linguistics (SEzik), 30*(1), 32–47.

Fielder, G. E. (1985), "Aspect and Modality in Bulgarian Subordinate Clauses" In M. S. Flier & A. Timberlake (Eds.), *UCLA Slav. Studies: 12. The Scope of Slavic Aspect* (pp. 181–193). Columbus, OH: Slavica.

Fielder, G. E. (1995), "Narrative Perspective and the Bulgarian L-Participle", *The Slavic and East European Journal, 39*(4), 585–600.

Filip, H. (1989), "Aspectual Properties of the AN-Construction in German" In W. Abraham & T. Janssen (Eds.), *Linguistische Arbeiten (LArb): 237. Tempus-Aspekt-Modus: Die lexikalischen und grammatischen Formen in den germanischen Sprachen* (pp. 259–292). Tübingen: Niemeyer.

Francis, W.N. & H. Kucera. (1979), *Manual of Information to Accompany a Standard Sample of Present-day Edited American English, for Use with Digital Computers.* Original ed. 1964, revised 1971, revised and augmented 1979. Providence, R.I.: Department of Linguistics, Brown University.

Francis, W.N. & H. Kucera. (1982), *Frequency Analysis of English Usage: Lexicon and Grammar.* Boston: Houghton Mifflin.

Fritz, S. (1982), "Ist das Bulgarische eine Aspekt- oder Tempussprache?", *Die Slawischen Sprachen, 1*, 24–32.

Gachelin, J.-M. (1997), "The Progressive and Habitual Aspects in Non-Standard Englishes" In E. W. Schneider & M. Görlach (Eds.), *Varieties of English around the worldgeneral series: Vol. 18. General studies, British Isles, North America* (pp. 33–46). Amsterdam: Benjamins.

Gass, S. (1996), "Second Language Acquisition and Linguistic Theory: the Role of Language Transfer" In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 317–345). SanDiego etc.: Academic Press.

Gass, S. M., & Selinker, L. (1994), *Second Language Acquisition: An Introductory Course*. Hillsdale NJ etc.: Erlbaum.

Gilquin, G. (2008), "Combining Contrastive and Interlanguage Analysis to Apprehend Transfer: Detection, Explanation, Evaluation" In G. Gilquin (Ed.), *Language and Computers: Vol. 66. Linking up Contrastive and Learner Corpus Research* (pp. 3–33). Amsterdam: Rodopi.

Gilquin, G., & Granger, S. (2011), "From EFL to ESL: Evidence from the International Corpus of Learner English" In J. Mukherjee & M. Hundt (Eds.), *Studies in Corpus Linguistics: Vol. 44. Exploring Second-Language Varieties of English and Learner Englishes. Bridging a Paradigm Gap* (pp. 55–78). Amsterdam: Benjamins.

Gilquin, G., & Paquot, M. (2007), "Spoken Features in Learner Academic Writing: Identification, Explanation and Solution" In M. Davies, P. Rayson, S. Hunston, & P. Danielsson (Eds.), *Proceedings of the Corpus Linguistics Conference 2007*. Birmingham.

Givon, T. (1987), "Beyond Foreground and Background" In R. S. Tomlin (Ed.), *Typological Studies in Language: Vol. 11. Coherence and Grounding in Discourse. Outcome of a Symposium. Eugene, Oregon* (pp. 175–188). Amsterdam etc.: Benjamins.

Givón, T. (Ed.) (1989), *Syntax and Semantics: Vol. 12. Discourse and Syntax* (3.th ed.). San Diego etc.: Acad. Press.

Godfrey, D. (1980), "A Discourse Analysis of Tense in Adult ESL Monologues" In D. Larsen-Freeman (Ed.), *Discourse Analysis in Second Language Research* (pp. 92–100). Rowley, MA: Newbury House Publ.

Goldschneider, J. M., & DeKeyser, R. M. (2001), "Explaining the 'Natural Order of L2 Morpheme Acquisition' in English: A Meta-Analysis of Multiple Determinants", *Language Learning: A Journal of Research in Language Studies, 51*(1), 1–50.

Granger, S. (1996), "From CA to CIA and back: an Integrated Approach to Computerized Bilingual and Learner Corpora" In K. Aijmer & B. Altenberg (Eds.), *Languages in Contrast. Text-Based Cross-Linguistic Studies* (pp. 37–51). Lund: Lund University Press.

Granger, S. (Ed.) (1998), *Learner English on Computer.* London etc.: Longman.

Granger, S. (1999), "Use of Tenses by Advanced EFL Learners: Evidence from an Error-Tagged Computer Corpus" In H. Hasselgård, S. Oksefiell, & S. Johansson (Eds.), *Language and Computers: Vol. 26. Out of Corpora: Studies in Honour of Stig Johansson* (pp. 191–202). Amsterdam: Rodopi.

Granger, S. (2002), "A Bird's-Eye View of Learner Corpus Research" In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3–33). Amsterdam, Netherlands: Benjamins.

Granger, S. (2003), "The Corpus Approach: a Common Way forward for Contrastive Linguistics and Translation Studies?" In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Approaches to Translation Studies (ApTS): 20. Corpus-Based Approaches to Contrastive Linguistics and Translation Studies* (pp. 17–29). Amsterdam, Netherlands: Rodopi.

Granger, S. (2004), "Computer Learner Corpus Research: Current Status and Future Prospects" In U. Connor & T. A. Upton (Eds.), *Language and Computers: Studies in Practical Linguistics*

(L&Comp): 52. Applied Corpus Linguistics: A Multidimensional Perspective (pp. 123–145). Amsterdam, Netherlands: Rodopi.

Granger, S. (2004), "Practical Applications of Learner Corpora" In B. Lewandowska-Tomaszczyk (Ed.), *Lodz Studies in Language: 9. Practical Applications in Language and Computers. PALC 2003* (pp. 291–301). Frankfurt, Germany: Peter Lang.

Granger, S. (2009), "The Contribution of Learner Corpora to Second Language Acquisition and Foreign Language Teaching" In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13–31). Amsterdam: Benjamins.

Granger, S., & Aarts, J. (1998), "Tag Sequences in Learner Corpora: a Key to Iinterlanguage Grammar and Discourse" In S. Granger (Ed.), *Learner English on Computer* (pp. 132–143). London etc.: Longman.

Granger, S., Dagneaux, E., & Meunier, F. (2002), *International Corpus of Learner English* (Version 1.1.). Louvain: Univ. Catholique Centre for English Corpus Linguistics.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009), *International Corpus of Learner English* (Version 2.). Louvain-la-Neuve UCL Presses Univ. de Louvain.

Granger, S., Hung, J., & Petch-Tyson, S. (Eds.) (2002), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching.* Amsterdam, Netherlands: Benjamins.

Granger, S., & Meunier, F. (1994), "Towards a Grammar Checker for Learners of English" In U. Fries, G. Tottie, & P. Schneider (Eds.), *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993* (pp. 79–93). Amsterdam, Netherlands; Atlanta, USA: Rodopi.

Granger, S., & Rayson, P. (1998), "Automatic Profiling of Learner Texts" In S. Granger (Ed.), *Learner English on Computer* (pp. 119–132). London etc.: Longman.

Granger, S., & Thewissen, J. (18-22.2005), "The Contribution of Error-Tagged Learner Corpora to the Assessment of Language Proficiency: Evidence from the International Corpus of Learner English", Paper presented at the 27th Language Testing Research Colloquium. Ottawa.

Granger, S., Gilquin, G. & Meunier, F. (2013), *Twenty Years of Learner Corpus Research: Looking back, Moving ahead. Corpora and Language in Use – Proceedings 1*, Louvain-la-Neuve, Presses universitaires de Louvain.

Halliday, M. A. K., & James, Z. (1993), "A Quantitative Study of Polarity and Primary Tense in the English Finite Clause" In J. M. Sinclair, M. Hoey, & G. Fox (Eds.), *Techniques of description. Spoken and written discourse* (pp. 32–66). London etc.: Routledge.

Hahn, A., Reich, S., & Schmied, J. (2000), "Aspect in the Chemnitz Internet Grammar" In C. Mair & M. Hundt (Eds.), *Language and computers: Vol. 33. Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20) Freiburg im Breisgau 1999* (pp. 131–139). Amsterdam [u.a.]: Rodopi.

Hardie, A. (2007), "Part-of-Speech Ratios in English Corpora", *International Journal of Corpus Linguistics, 12*(1), 55–81.

Haznedar, B. (2007), "The Acquisition of Tense-Aspect in Child Second Language English", *Second Language Research, 23*(4), 383–417.

Hewings, A., & Hewings, M. (2005), *Grammar and Context: An Advanced Resource Book*. London etc.: Routledge.

Hinkel, E. (1992), "L2 Tense and Time Reference", *TESOL Quarterly, 26*(3), 557–572.

Hinkel, E. (1997), "The Past Tense and Temporal Verb Meanings in a Contextual Frame", *TESOL Quarterly, 31*(2), 289–313.

Hinkel, E. (2002), *Second Language Writers' Text: Linguistic and Rhetorical Features. ESL and Applied Linguistics Professional Series.* Mahwah, NJ: Lawrence Erlbaum.

Hinkel, E. (2002), "Teaching Grammar in Writing Classes: Tenses and Cohesion" In E. Hinkel & S. Fotos (Eds.), *ESL and Applied Linguistics Professional Series. New Perspectives on Grammar Teaching in Second Language Classrooms* (pp. 181–198). Mahwah, N.J: Erlbaum Assoc.

Hinkel, E. (2004), "Tense, Aspect and the Passive Voice in L1 and L2 Academic Texts", *Language Teaching Research, 8*(1), 5–29.

Hopper, P. J. (1989), "Aspect and Foregrounding in Discourse" In T. Givón (Ed.), *Syntax and Semantics: Vol. 12. Discourse and Syntax* (3rd ed., pp. 213–241). San Diego etc.: Acad. Press.

Housen, A. (2000), "Verb Semantics and the Acquisition of Tense-Aspect in L2 English", *Studia Linguistica, 54*(2), 249–259.

Housen, A. (2002a), "A Corpus-Based Study of the L2-Acquisition of the English Verb System", In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 77--119). Amsterdam, Netherlands: Benjamins.

Housen, A. (2002b), "The Development of Tense-Aspect in English as a Second Language and the Variable Influence of Inherent Aspect" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 155–197). Amsterdam etc.: Benjamins.

Huddleston, R. D. (1984), *Introduction to the Grammar of English.* Cambridge etc: Cambridge Univ. Press.

Huddleston, R. D., Pullum, G. K., & Bauer, L. (2005), *The Cambridge Grammar of the English Language* (3. print.). Cambridge: Cambridge Univ. Press.

Hundt, M. (2004), "Animacy, Agentivity, and the Spread of the Progressive in Modern English", *English Language and Linguistics, 8*(1), 47–69.

Hundt, M. (2004), "The Passival and the Progressive Passive: A Case Study of Layering in the English Aspect and Voice Systems" In H. Lindquist & C. Mair (Eds.), *Studies in Corpus Linguistics: Vol. 13. Corpus Approaches to Grammaticalization in English* (pp. 79–121). Amsterdam: Benjamins.

Hundt, M. (2009), "Global English - Global Corpora: Report on a Panel Discussion at the 28th ICAME Conference" In A. Renouf (Ed.), *Language and computers: Vol. 69. Corpus Linguistics. Refinements and Reassessments* (pp. 451–462). Amsterdam etc.: Rodopi.

Hundt, M. (2009), "Global Feature - Local Norms?: A Case Study on the Progressive Passive" In T. Hoffmann & L. Siebers (Eds.), *Selected Papers from the 13th IAWE Conference: Vol. 13. World Englishes - Problems, Properties and Prospects* (pp. 287–308). Amsterdam etc.: Benjamins.

Hundt, M., & Mukherjee, J. (2011), "Discussion Forum: New Englishes and Learner Englishes - Quo Vadis?" In J. Mukherjee & M. Hundt (Eds.), *Studies in Corpus Linguistics: Vol. 44. Exploring Second-Language Varieties of English and Learner Englishes. Bridging a Paradigm Gap* (pp. 209–217). Amsterdam: Benjamins.

Hundt, M., Sand, A., & Siemund, R. (1998), *Manual of Information to Accompany The Freiburg - LOB Corpus of British English (FLOB),* Albert-Ludwigs-Universität Freiburg: http://khnt.hit.uib.no/icame/manuals/*FLOB*/index.htm.

Hundt, M., Sand, A., & Skandera, P. (1999), *Manual of Information to Accompany the Freiburg - Brown Corpus of American English (FROWN),* Albert-Ludwigs-Universität Freiburg: http://khnt.hit.uib.no/icame/manuals/*FROWN*/INDEX.HTM.

Hundt, M., & Smith, N. (2009), "The Present Perfect in British and American English: has there been any Change, Recently?", *ICAME Journal, 33*, 45–63.

Hundt, M., & Vogel, K. (2011), "Overuse of the Progressive in ESL and Learner Englishes - Fact or Fiction?" In J. Mukherjee & M. Hundt (Eds.), *Studies in Corpus Linguistics: Vol. 44. Exploring Second-Language Varieties of English and Learner Englishes. Bridging a Paradigm Gap* (pp. 145–165). Amsterdam: Benjamins.

Hunston, S. (2002), *Corpora in Applied Linguistics* (repr.). Cambridgeetc.: Cambridge Univ. Press.

Izquierdo, J., & Collins, L. (2008), "The Facilitative Role of L1 Influence in Tense-Aspect Marking: a Comparison of Hispanophone and Anglophone Learners of French", *The Modern Language Journal, 92,* 350-368.

Jappy, T. (1996), "Investigating Grounding Across Narrative and Oral Discourse" In S. Botley, J. Glass, T. McEnery, & A. Wilson (Eds.), *UCREL Technical Papers. Proceedings of Teaching and Language Corpora 1996.* TALC 96 (pp. 138–149). Lancaster: UCREL.

Kabakciev, K. (1989), "On 'Telicity' and Related Problems" In L.-G. Larsson (Ed.), *Acta Universitatis Upsaliensis: Studia Uralica et Altaica Upsaliensia: 19. Proceedings of the Second Scandinavian Symposium on Aspectology* (pp. 13–32). Uppsala: Almqvist & Wiksell.

Kachru, B. B. (2008), *The Handbook of World Englishes* (2nd print.). Malden, Mass.: Blackwell.

Kaplan, J., & Shand, M. (1984), "Error Detection as a Function of Integrativity" In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of Second Language Acquisition* (pp. 51–59). Rowley, Mass.: Newbury House Publ.

Kellerman, E. (1997), "Why Typologically Close Languages are Interesting for the Study of Second Language Acquisition" In J. Aarts, I. de Mönnink, & H. Wekker (Eds.), *Studies in English Language and Teaching: in Honour of Flor Aarts* (pp. 275–293). Amsterdam, Netherlands; Atlanta, USA: Rodopi.

Klein, W. (1992), "The Present Perfect Puzzle", *Language: Journal of the Linguistic Society of America, 68*(3), 525–552.

Klein, W. (1994), *Time in Language* (1. publ). London: Routledge.

Klein, W. (1995), "A Simplest Analysis of the English Tense-Aspect System" In W. Riehle (Ed.), *Proceedings of the Conference of the German Association of University Teachers of English: Vol. 16. Proceedings of the Anglistentag 1994, Graz* (pp. 139–151). Tübingen: Niemeyer.

Klein, W. (2000), "An Analysis of the German Perfekt", *Language: Journal of the Linguistic Society of America, 76*(2), 358–382.

Klein, W. (2009), "Concepts of Time" In W. Klein & P. Li (Eds.), *The Expression of Cognitive Categories: Vol. 3. The Expression of Time* (pp. 5–38). Berlin: Mouton de Gruyter.

Klein, W. (2009), "How Time is Encoded" In W. Klein & P. Li (Eds.), *The Expression of Cognitive Categories: Vol. 3. The Expression of Time* (pp. 39–81). Berlin: Mouton de Gruyter.

Klein, W., & Li, P. (Eds.) (2009), *The Expression of Cognitive Categories: Vol. 3: The Expression of Time.* Berlin: Mouton de Gruyter.

Klein, W., & Vater, H. (1998), "The Perfect in English and German", In L. I. Kulikov & H. Vater (Eds.), *Linguistische Arbeiten: Vol. 382. Typology of verbal categories. Papers Presented to Vladimir Nedjalkov on the Occasion of his 70th Birthday* (pp. 215–237). Tübingen: Niemeyer.

Koch, P.,& W. Oesterreicher (1985), *Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte*. In: Romanistisches Jahrbuch 36. Berlin etc.: Walter de Gruyter, 15-43.

König, E. (1995), "On Analyzing the Tense-Aspect System of English: a State-Of-The-Art Report" In W. Riehle (Ed.), *Proceedings of the Conference of the German Association of University Teachers of English: Vol. 16. Proceedings of the Anglistentag 1994, Graz* (pp. 153–169). Tübingen: Niemeyer.

König, E., & Gast, V. (2009), *Understanding English-German Contrasts. ESV basics: Vol. 29.* Berlin: Schmidt.

Königs, K. (1995), "Zur Übersetzung der englischen Verlaufsform ins Deutsche", *Lebende Sprachen: Zeitschrift für Fremde Sprachen in Wissenschaft und Praxis, 40*(4), 153–158.

Kortmann, B. (1995), "Compositionality and the Perfect" In W. Riehle (Ed.), *Proceedings of the Conference of the German Association of University Teachers of English: Vol. 16. Proceedings of the Anglistentag 1994, Graz* (pp. 183–199). Tübingen: Niemeyer.

Krashen, S. D. (1978), "Is the "Natural Order" an Artifact of the Bilingual Syntax Measure?: Comments on J. Porter's "A Cross-Sectional Study of Morpheme Acquisition in First Language Learners". *Language Learning, 28*(1), 187–191.

Lado, R. (1957), *Linguistics Across Cultures: Applied Linguistics for Language Teachers.* Ann Arbor: Univ. of Michigan Press.

Leech, G. N. (1987), *Meaning and the English Verb* (2. ed.). London etc.: Longman.

Leech, G. (1998), "Preface" In S. Granger (Ed.), *Learner English on Computer* (pp. xiv–xxi). London etc.: Longman.

Leech, G., & Smith, N. (2006), "Recent Grammatical Change in Written English 1961-1992: Some Preliminary Findings of a Comparison of American with British English" In A. Renouf & A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics* (pp. 185–204). Amsterdam etc.: Rodopi.

Leech, G., & Smith, N. (2009), "Change and Constancy in Linguistic Change: How Grammatical Usage in Written English Evolved in the Period 1931-1991" In A. Renouf (Ed.), *Language and Computers: Vol. 69. Corpus Linguistics. Refinements and Reassessments* (pp. 173–200). Amsterdam etc.: Rodopi.

Lenko-Szymanska, A. (2007), "Past Progressive or Simple Past?: The Acquisition of Progressive Aspect by Polish Advanced Learners of English" In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Language and Computers: Vol. 61. Corpora in the Foreign Language Classroom. Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6) University of Granada Spain 4 - 7 July 2004* (pp. 253–267). Amsterdam etc.: Rodopi.

Lennon, P. (1991), "Error: Some Problems of Definition Identification and Distinction", *Applied Linguistics, 12*(2), 180–196.

Li, P., & Shirai, Y. (2000), *The Acquisition of Lexical and Grammatical Aspect. Studies on Language Acquisition: Vol. 16.* Berlin etc.: Mouton de Gruyter.

Li, P., & Zhao, X. (2009), "Computational Modeling of the Expression of Time" In W. Klein & P. Li (Eds.), *The Expression of Cognitive Categories: Vol. 3. The Expression of Time* (pp. 241–271). Berlin: Mouton de Gruyter.

Lindstedt, J. (1985), *On the Semantics of Tense and Aspect in Bulgarian. Slavica Helsingiensia: Vol. 4.* Helsinki: Univ. Press.

Lindstedt, J. (2000), "The Perfect-Aspectual, Temporal and Evidential" In Ö. Dahl (Ed.), *Empirical Approaches to Language Typology (EALT): Eurotyp 20-6. Tense and Aspect in the Languages of Europe* (pp. 365–383). Berlin, Germany: Mouton de Gruyter.

Löbner, S. (2002), "Is the German Perfekt a Perfect Perfect?" In I. Kaufmann & B. Stiebels (Eds.), *Studia Grammatica (Stgr): 53. More Than Words* (pp. 369–391). Berlin, Germany: Akademie.

Lorenz, G. (2002), "Language Corpora Rock the Base: On Standard English Grammar, Perfective Aspect and Seemingly Adverse Corpus Evidence" In B. Kettemann, G. Marko, & T. McEnery (Eds.), *Language and Computers: Vol. 42. Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July 2000* (pp. 131–145). Amsterdam etc.: Rodopi.

Lüdeling, A., & Kytö, M. (Eds.) (2008), *Corpus Linguistics: An International Handbook*. Volume 1. Berlin: de Gruyter.

Lüdeling, A., & Kytö, M. (Eds.) (2009), *Corpus Linguistics: An International Handbook*. Volume 2. Berlin: de Gruyter.

Mair, C. (1997), "Parallel Corpora: a Real-Time Approach to the Study of Language Change in Progress" In M. Ljung (Ed.), *Corpus-Based Studies in English. Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17), Stockholm, May 15-19, 1996* (pp. 195–209). Amsterdam etc.: Rodopi.

Mair, C., & M. Hundt (1995), "Why is the Progressive becoming More Frequent in English?: A Corpus-Based Investigation of language Change in Progress", *Zeitschrift für Anglistik und Amerikanistik, 43*(2), 111–122.

Markus, M. (1977), *Tempus und Aspekt: Zur Funktion von Präsens, Präteritum und Perfekt im Englischen und Deutschen. Kritische Information: Vol. 61*. München: Fink.

McEnery, T., & Wilson, A. (1996), *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

McEnery, A., Xiao, Z., & Mo, L. (2003), "Aspect Marking in English and Chinese: Using the Lancaster Corpus of Mandarin Chinese for Contrastive Language Study", *Literary & Linguistic Computing, 18*(4), 361–378.

McEnery, T., Tono, Y., & Xiao, R. (2006), *Corpus-Based Language Studies: An Advanced Resource Book*. London etc.: Routledge.

Meunier, F. (1998), "Computer Tools for the Analysis of Learner Corpora" In S. Granger (Ed.), *Learner English on Computer* (pp. 19–39). London etc.: Longman.

Meunier, F. (2002), "The Pedagogical Value of Native and Learner Corpora in EFL Grammar Teaching" In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Language Learning & Language Teaching: 6. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 119–141). Amsterdam, Netherlands: Benjamins.

Mindt, D. (1997), "Corpora and the Teaching of English in Germany" In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 40–50). London [u.a]: Longman.

Mindt, D. (2000), *An Empirical Grammar of the English Verb System* (1. Aufl.). Berlin: Cornelsen.

Mindt, D. (2002), "A Corpus-Based Grammar for ELT" In B. Kettemann, G. Marko, & T. McEnery (Eds.), *Language and Computers: Vol. 42. Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July 2000* (pp. 91–105). Amsterdam etc.: Rodopi.

Mukherjee, J. (2006), "Corpus Linguistics and English Reference Grammars" In A. Renouf & A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics* (pp. 337–354). Amsterdam etc.: Rodopi.

Mukherjee, J. (2006), "Corpus Linguistics and Language Pedagogy: the State of the Art - and beyond" In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *English Corpus Linguistics: Vol. 3. Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods* (pp. 5–24). Frankfurt am Main: Lang.

Mukherjee, J. (2006), "Rethinking Applied Corpus Linguistics from a Language-Pedagogical Perspective: New Departures in Learner Corpus Research" In B. Kettemann (Ed.), *Planning, Gluing and Painting Corpora. Inside the Applied Corpus Linguist's Workshop* (pp. 205–232). Frankfurt am Main: Peter Lang.

Nesselhauf, N. (2004), "Learner Corpora and their Potential for Language Teaching" In J. M. Sinclair (Ed.), *Studies in Corpus Linguistics: Vol. 12. How to Use Corpora in Language Teaching* (pp. 125–152). Amsterdam: Benjamins.

Nesselhauf, N. (2006), "Researching L2 Production with *ICLE*" In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *English Corpus Linguistics: Vol. 3. Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods* (pp. 141–156). Frankfurt am Main: Lang.

Nesselhauf, N. (2007), "The Spread of the Progressive and Its 'Future' Use", *English Language and Linguistics, 11*(1), 191–207.

Nesselhauf, N., & Römer, U. (2007), "Lexical-Grammatical Patterns in Spoken English: The Case of the Progressive with Future Time Reference", *International Journal of Corpus Linguistics, 12*(3), 297–333.

Noyau, C. (2002), "Temporal Relations in Learner Varieties: Grammaticalization and Discourse Construction" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 107–127). Amsterdam etc.: Benjamins.

Odlin, T. (1989), *Language Transfer: Cross-Linguistic Influence in Language Learning. The Cambridge Applied Linguistics Series.* Cambridge: Cambridge Univ. Press.

Odlin, T., Alonso, R. A., & Alonso-Vazquez, C. (2006), "Fossilization in L2 and L3" In Z. Han & T. Odlin (Eds.), *Second Language Acquisition: Vol. 14. Studies of Fossilization in Second Language Acquisition* (pp. 83–99). Clevedon, U.K., Buffalo: Multilingual Matters.

Odlin, T. (2008), "Conceptual Transfer and Meaning Extensions" In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 306–341). New York etc.: Routledge.

Pica, T. (1983), "Adult Acquisition of English as a Second Language Under Different Conditions of Exposure", *Language Learning, 33*(4), 465–497.

Pica, T. (1983), "Methods of Morpheme Quantification: Their Effect on the Interpretation of Second Language Data", *Studies in Second Language Acquisition, 6*(1), 69–78.

Pica, T. (1984), "L1 Transfer and L2 Complexity as Factors in Syllabus Design", *TESOL Quarterly, 18*(4), 689–704.

Pica, T. (2008), "SLA in the Instructional Environment". *Working Papers in Educational Linguistics*, 23(1), 1-27.

Pienemann, M. (1987), "Psychological Constraints on the Teachability of Languages". In C. W. Pfaff (Ed.), *Cross-Linguistic Series on Second Language Research. First and Second Language Acquisition Processes* (pp. 143–168). Cambridge, Mass.: Newbury House Publ.

Pienemann, M. (1999). *Language Processing and Second Language Development: Processability Theory. Studies in Bilingualism: Vol. 15.* Amsterdam: Benjamins.Pravec, N. A. (2002), "Survey of Learner Corpora", *ICAME Journal, 26*, 81–114.

Quirk, R. (1985), *A Comprehensive Grammar of the English Language.* London etc.: Longman.

Quirk, R., Leech, G., & Svartvik, J. (1974), *A Grammar of Contemporary English.* London: Longman.

Radden, G., & Dirven, R. (2007). *Cognitive English Grammar*. Amsterdam etc.: Benjamins.

Radeva, V. (2003), *Bulgarische Grammatik: Morphologisch-syntaktische Grundzüge.* Hamburg: Buske.

Radford, A. (1988), *Transformational Grammar: A First Course. Cambridge*: Cambridge Univ.Press.

Ranta, E. (2006), "The 'Attractive' Progressive: Why Use the -ing Form in English as a Lingua Franca?", *NJES: Nordic Journal of English Studies, 5*(2), 95–116.

Rayson, P., & Garside, R. (2000), "Comparing Corpora Using Frequency Profiling" In *Proceedings of the Workshop on Comparing Corpora*, (pp. 1–6). Hong Kong.

Rayson, P., Berridge, D., & Francis, B. (2004), "Extending the Cochran Rule for the Comparison of Word Frequencies between Corpora" In G. Purnelle & C. D. A. Fairon (Eds.), *Le Pids des Mots: Proceedings of the 7th International Conference on Statistical Analysis of Textual Data*. JADT 2004 (pp. 926–936). Louvain-la-Neuve: Presses universitaires de Louvain.

Rayson, P. (2007), *Wmatrix: a Web-Based Corpus Processing Environment*. Lancaster: Computing Department, Lancaster University, from http://www.comp.lancs.ac.uk/ucrel/wmatrix/

Reichenbach, H. (1952). *Elements of Symbolic Logic* (4. print.). New York: Macmillan.

Robison, R. E. (1990), "The Primacy of Aspect", *Studies in Second Language Acquisition, 12*(03), 315–330.

Robison, R. (1995), "The Aspect Hypothesis Revisited: A Cross-Sectional Study of Tense and Aspect Marking in Interlanguage", *Applied Linguistics, 16*(3), 344–370.

Rohde, A. (2002), "The Aspect Hypothesis in Naturalistic L2 Acquisition: What Uninflected and Non-Target-Like Verb Forms in Early Interlanguage Tell us" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 199–220). Amsterdam etc.: Benjamins.

Römer, U. (2004), "Comparing Real and Ideal Language Learner Input: The Use of EFL Textbook Corpus in Corpus Linguistics and Language Teaching" In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Studies in Corpus Linguistics: Vol. 17. Corpora and Language Learners* (pp. 151–169). Amsterdam etc.: Benjamins.

Römer, U. (2005), *Progressives, Patterns, Pedagogy.: A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts, and Didactics.* Amsterdam, Philadelphia: J. Benjamins.

Römer, U. (2006), "Looking at Looking: Functions and Contexts of Progressives in Spoken English and 'School' English" In A. Renouf & A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics* (pp. 231–242). Amsterdam etc.: Rodopi.

Römer, U. (2008), "Corpora and Language Teaching" In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook. Volume 1* (pp. 112–130). Berlin: de Gruyter.

Salaberry, M. R. (1999), "The Development of Past Tense Verbal Morphology in Classroom L2 Spanish", *Appied Linguistics, 20*(2), 151–178.

Salaberry, M. R. (2000), "The Acquisition of English Past Tense in an Instructional Setting", *System, 28*(1), 135–152.

Salaberry, M. R. (Ed.) (2013), *Studies in Second and Foreign Language Education: Vol. 2. Research Design and Methodology in Studies on L2 Tense and Aspect.* Boston etc.: de Gruyter Mouton.

Salaberry, R., & Yasuhiro Shirai (2002), "L2 Acquisition of Tense-Aspect Morphology" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 1–19). Amsterdam etc.: Benjamins.

Salaberry, M. & Comajoan, L. (Eds.) (2013), *Research Design and Methodology in Studies on L2 Tense and Aspect.* Berlin etc: De Gruyter.

Saville-Troike, M. (2006), *Introducing Second Language Acquisition.* Cambridge: Cambridge University Press.

Scatton, E. A. (1984), *A Reference Grammar of Modern Bulgarian.* Columbus O.: Slavica Publ.

Scatton, E. A. (2002), "Bulgarian" In B. Comrie (Ed.), *The Slavonic Languages* (pp. 188–249). London etc.: Routledge.

Schlüter, N. (2000), "The Present Perfect in British and American English: Selected Results of an Empirical Analysis" In C. Mair & M. Hundt (Eds.), *Language and Computers: Vol. 33. Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20) Freiburg im Breisgau 1999* (pp. 313–320). Amsterdam etc.: Rodopi.

Schlüter, N. (2002), *Present Perfect: Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht. Language in Performance.* Tübingen: Narr.

Schlüter, N. (2002), "Temporal Specification of the Present Perfect: A Corpus-Based Study" In P. Peters, P. Collins, & A. Smith (Eds.), *Language and Computers: Studies in Practical Linguistics (L&Comp): 36. New Frontiers of Corpus Research* (pp. 307–315). Amsterdam, Netherlands: Rodopi.

Schlüter, N. (2006), "How Reliable are the Results?: Comparing Corpus-Based Studies of the Present Perfect", *Zeitschrift für Anglistik und Amerikanistik, 54*(2), 135–148.

Schüller, S. (2005), "Problematik der Darstellung von Aspektunterschieden in Englischen Referenzgrammatiken" In T. Herbst (Ed.), *Linguistische Dimensionen des Fremdsprachenunterrichts* (pp. 29–42). Würzburg: Königshausen und Neumann.

Scott, M. (2004), *Wordsmith Tools Version 4.0.* Oxford: Oxford University Press.

Selinker, L. (1972), "Interlanguage", *International Review of Applied Linguistics in Language Teaching, 10*, 209–231.

Selinker, L. (1976), "Interlanguage" In J. H. Schumann & N. Stenson (Eds.), *New Frontiers in Second Language Learning* (pp. 114–136). Rowley, Mass.: Newbury House.

Sharwood Smith, M. (1994), Second Language Learning: Theoretical Foundations. London etc.: Longman.

Shirai, Y. (1998), "The Emergence of Tense-Aspect Morphology in Japanese: Universal Predisposition?", *First Language (FirstL), 18*(3 [54]), 281–309.

Shirai, Y., & Nishi, Y. (2003), "Lexicalisation of Aspectual Structures in English and Japanese" In A. Giacalone Ramat (Ed.), Empirical Approaches to Language Typology. Berlin etc.: Mouton de Gruyter, 267–290.

Shirai, Y. (2007), "The Aspect Hypothesis, the Comparative Fallacy and the Validity of Obligatory Context Analysis: a Reply to Lardiere, 2003", *Second Language Research, 23*(1), 51–64.

Shirai, Y. (2009), "Temporality in First and Second Language Acquisition" In W. Klein & P. Li (Eds.), *The Expression of Cognitive Categories: Vol. 3. The Expression of Time* (pp. 167–193). Berlin: Mouton de Gruyter.

Shirai, Y. (2010), "Semantic bias and morphological regularity in the acquisition of tense-aspect morphology: what is the relation?", *Linguistics* 48(1), 171-194.

Shirai, Y., & Andersen, R. W. (1995), "The Acquisition of Tense-Aspect Morphology: a Prototype Account", *Language: Journal of the Linguistic Society of America, 71*(4), 743–762.

Slabakova, R. (2000), "L1 Transfer Revisited: The L2 Acquisition of Telicity Marking in English by Spanish and Bulgarian Native Speakers", *Linguistics: An Interdisciplinary Journal of the Language Sciences, 38*(4), 739–770.

Slabakova, R. (2002), "Recent Research on the Acquisition of Aspect: an Embarrassment of Riches?", *Second Language Research, 18*(2), 172–188.

Slabakova, R., & Silvina Montrul (2002), "On Viewpoint Aspect Interpretation and its L2 Acquisition" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 363–395). Amsterdam etc.: Benjamins.

Smith, C. S. (1983), "A Theory of Aspectual Choice", *Language, 59*(3), 479–501.

Smith, N., & Rayson, P. (2007), "Recent Change and Variation in the British English Use of the Progressive Passive", *ICAME Journal, 31*, 129–159.

Smitterberg, E. (2000), "The Progressive Form and Genre Variation During the Nineteenth Century" In R. Bermúdez-Otero, D. Denison, R. M. Hogg, & C. B. McCully (Eds.), *Topics in English linguistics: Vol. 31. Generative Theory and Corpus Studies. A Dialogue from 10 ICEHL* (pp. 283–297). Berlin etc.: Mouton de Gruyter.

Smitterberg, E. (2005), *The Progressive in 19th-Century English: A Process of Integration. Language and Computers: Vol. 54.* Amsterdam etc.: Rodopi.

Smitterberg, E. (2008), "The Progressive and Phrasal Verbs: Evidence of Colloquialization in Nineteenth-Century English?" In T. Nevalainen, I. Taavitsainen, P. Pahta, & M. Korhonen (Eds.), *Studies in Language Variation: Vol. 2. The Dynamics of Linguistic Variation. Corpus Evidence on English Past and Present* (pp. 269–290). Amsterdam etc.: Benjamins.

Smitterberg, E., Reich, S., & Hahn, A. (2000), "The Present Progressive in Political and Academic Language in the 19th and 20th Centuries: a Corpus-Based Investigation", *ICAME Journal, 24*, 99–118.

Stechow, A. von (2002), "German seit 'Since' and the Ambiguity of the German Perfect" In I. Kaufmann & B. Stiebels (Eds.), Studia Grammatica (Stgr): 53. More Than Words (pp. 393–432). Berlin, Germany: Akademie.

Stutterheim, C. von, & Klein, W. (1987), "A Concept-Oriented Approach to Second Language Studies" In C. W. Pfaff (Ed.), *Cross-Linguistic Series on Second Language Research. First and Second Language Acquisition Processes* (pp. 191–205). Cambridge, Mass.: Newbury House Publ.

Sugaya, N., & Shirai, Y. (2007), "The Acquisition of Progressive and Resultative Meanings of the Imperfective Aspect Marker by L2 Learners of Japanese: Transfer, Universals, or Multiple Factors?", *Studies in Second Language Acquisition, 29*, 1–38.

Swan, M. (1987), "German Speakers" In M. Swan & B. Smith (Eds.), *Learner English. A Teacher's Guide to Interference and other Problems* (pp. 30–42). Cambridge etc.: Univ. Press.

Swan, M. (2004), "Complex Quantum Chaos and the Present Perfect", *ELT Journal, 58*(1), 68–70.

Swan, M., & Smith, B. (Eds.) (1987), *Learner English: A Teacher's Guide to Interference and other Problems.* Cambridge etc.: Univ. Press.

Tono, Y. (2000), "A Computer Learner Corpus Based Analysis of the Acquisition Order of English Grammatical Morphemes" In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective. Papers from the Third International Conference on Teaching and Language Corpora* (pp. 123–132). Lang, Peter Frankfurt.

Tono, Y. (2003), "Learner Corpora: Design, Development and Applications" In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Technical Papers. Proceedings of the Corpus Linguistics 2003 Conference* (pp. 800–809). Lancaster: UCREL.

van Rooy, B. (2006), "The Extension of the Progressive Aspect in Black South African English", *World Englishes: Journal of English as an International and Intranational Language (WEng), 25*(1), 37–64.

van Rooy, B. (2008), "An Alternative Interpretation of Tense and Aspect in Black South African English", *World Englishes: Journal of English as an International and Intranational Language (WEng), 27*(3/4), 335–358.

van Rooy, B. (2009), "The Shared Core of the Perfect across Englishes: a Corpus-Based Analysis" In T. Hoffmann & L. Siebers (Eds.), *Selected Papers from the 13th IAWE Conference: Vol. 13. World Englishes - Problems, Properties and Prospects* (pp. 309–330). Amsterdam etc.: Benjamins.

van Rooy, B. (2011), "A Principled Distinction between Error and Conventionalized Innovation in African Englishes" In J. Mukherjee & M. Hundt (Eds.), *Studies in Corpus Linguistics: Vol. 44. Exploring Second-Language Varieties of English and Learner Englishes. Bridging a Paradigm Gap.* Amsterdam: Benjamins.

VanPatten, B. (1984), "Processing Strategies and Morpheme Acquisition" In F. R. Eckman, L. H. Bell, & D. Nelson (Eds.), *Universals of Second Language Acquisition* (pp. 88–98). Rowley, Mass.: Newbury House Publ.

Vendler, Z. (1957), "Verbs and Times", *Philosophical Review, 66*, 143–160.

Verkuyl, H. J. (2001), "Aspect in English: A Bulgarian Perspective", *Journal of Semantics: An International Journal for the Interdisciplinary Study of the Semantics of Natural Language, 18*(4), 367–390.

Verkuyl, H. J. (2005), "How (In)sensitive is Tense to Aspectual Information?" In B. Hollebrandse (Ed.), *Cahiers Chronos: Vol. 13. Crosslinguistic Views on Tense, Aspect and Modality* (pp. 145–169). Amsterdam etc.: Rodopi.

Virtanen, T. (1997), "The Progressive in NS and NNS Student Compositions: Evidence from the International Corpus of Learner English" In M. Ljung (Ed.), *Corpus-Based Studies in English. Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17), Stockholm, May 15-19, 1996* (pp. 299–309). Amsterdam etc.: Rodopi.

Voigt, K. (2005), "The Reliability of Temporal Adverbials as Trigger Words for Learning the English Present Perfect: Already, Always, and For" In *Discourse and Interaction I. Proceedings of the Brno Seminar on Linguistic Studies in English.* (pp. 121–130). Brno: Masaryk University Press.

Weist, R. M. (2002), "The First Language Acquisition of Tense and Aspect: a Review" In R. Salaberry, Y. Shirai, D. Li, R. M. Weist, R. W. Andersen, C. Noyau, et al. (Eds.), *Language Acquisition & Language Disorders: Vol. 27. The L2 Acquisition of Tense-Aspect Morphology* (pp. 21–78). Amsterdam etc.: Benjamins.

Wible, D., & Huang, P.-Y. (2003), "Using Learner Corpora to Examine L2 Acquisition of Tense-Aspect Morphology" In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Technical Papers. Proceedings of the Corpus Linguistics 2003 Conference* (pp. 889–898). Lancaster: UCREL.

Williams, M. S. (1990), "Perfect and Reported Forms in Bulgarian: Semantic Distinctions and Similarities", *ZB (Zeitschrift für Balkanologie), 26*(2), 227–235.

Wulff, S., Ellis, N. C., Romer, U., Bardovi Harlig, K., & Leblanc, C. J. (2009), "The Acquisition of Tense-Aspect: Converging Evidence from Corpora and Telicity Ratings", *Modern Language Journal (MLJ), 93*(3), 354–369.

Wulff, S., & Römer, U. (2009), "Becoming a Proficient Academic Writer: Shifting Lexical Preferences in the Use of the Progressive", *Corpora, 4*(2), 115–133.

Xiao, R. Z., & McEnery, T. (2005), "Situation Aspect: A Two-Level Approach" In B. Hollebrandse (Ed.), *Cahiers chronos: Vol. 13. Crosslinguistic Views on Tense, Aspect and Modality* (pp. 185–200). Amsterdam etc.: Rodopi.

Yang, S., & Huang, Y. Y. (2004), "The Impact of the Absence of Grammatical Tense in L1 on the Acquisition of the Tense-Aspect System in L2", *International Review of Applied Linguistics in Language Teaching, 42*(1), 49–70.

Zydatiss, W. (1977), "Die Bestimmung der Interimsgrammatik von Fremdsprachenlernern mit Hilfe von Fehleranalyse und Elicitation Procedures dargestellt am Beispiel der Progressive Form und des Present Perfect", *Die Neueren Sprachen, 26*, 14–37.

Appendix

A1 Tables and Frequencies

A1.1 Verb Tag and Verb Phrase Frequencies Six Corpora

	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
finite verb tags	22377	24571	8068	16457	8567	9721
finite verb tags %	11,8	11,4	10,8	11,6	8,2	8,5
Total POS tags	189934	214954	74627	142020	104250	114948

verb forms	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
simple present	15272	14303	4228	9310	4198	5238
simple past	1726	4462	802	2632	2368	2578
present progressive	320	414	165	464	175	133
past progressive	30	150	18	87	62	54
perfect progressive	35	39	17	17	19	14
present perfect	1112	807	632	901	423	362
past perfect	88	324	23	97	159	142
modal / future VPs	3798	4079	2178	2926	1178	1194
"going to" future	31	32	22	40	8	20
TOTAL	22412	24610	8085	16474	8590	9735

A1.2 Verb Types, Tokens and Lexical Verb Types in the Progressive and the Perfect

	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
Types Perfect	266	262	168	249	208	216
Tokens Perfect	1200	1131	656	998	581	504
Types Progressive	154	195	97	192	134	113
Tokens Progressive	385	603	200	192	256	198

Perfect VPs	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
activity	178	151	93	159	102	97
state	272	257	171	267	137	109
accomplishment	312	237	181	230	108	107
achievement	437	486	210	339	232	187
Total	1199	1131	655	995	579	500

Progressive VPs	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
activity	176	348	77	256	108	95
state	49	99	14	55	21	13
accomplishment	76	67	41	101	58	43
achievement	84	89	68	156	69	50
Total	385	603	200	568	256	201

A 1.3 Progressive and Perfect VPs in Main and Subordinate Clauses

progressive	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
main	210	324	107	295	150	121
subordinate	175	279	93	273	106	80
Total	385	603	200	568	256	201

perfect	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
main	641	586	461	654	362	293
subordinate	559	545	194	344	219	211
Total	1200	1131	655	998	581	504

A 1.4 Contracted Auxiliaries with the Progressive and the Perfect and Temporal Adverb Frequencies with the Perfect

	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
contracted aux. prog. ratio	20/384	66/603	0	17/559	6/257	23/202
contracted aux. perf. ratio	34/1200	61/1131	3/658	24/1000	14/581	33/516

Temporal Adverbials with the Perfect in %	BUCLE	GICLE	LOCNESS_br	LOCNESS_us	FLOB_F	FROWN_F
always	23,2	8,7	10,3	3,4	3,3	6,4
already	6,7	8,5	13,3	3,4	7,5	5,6
ever	8,1	5,6	1,2	3,4	0,0	4,8
never	8,1	10,0	5,5	5,3	6,1	11,2
since + NP	3,9	7,2	9,7	11,4	6,1	8,8
for + NP	7,3	9,7	12,1	14,4	13,6	12,8
just	2,8	4,9	0,6	2,7	0,9	3,2
recently	1,4	2,1	7,3	3,0	1,9	3,2
in + NP	4,2	3,8	9,7	12,2	10,8	4,0
over + NP	1,4	0,8	8,5	6,8	3,8	4,0
now(adays)	3,1	1,5	5,5	3,4	1,9	0,8
during + NP	2,2	3,3	1,2	3,4	3,3	0,8
throughout + NP	3,9	0,5	1,2	3,0	0,9	0,8
NP ago	1,4	3,1	0,0	0,4	0,0	0,8
long	1,4	1,0	0,6	0,4	0,9	4,8
yet	0,8	3,6	3,0	1,1	2,3	0,8
other adverbials	19,9	25,6	10,3	22,1	36,6	27,2
TOTAL	357	390	165	263	213	125

A 1.5 Essay codes in the error-tagged subcorpora *BUCLE*_110,000 and *GICLE*_110,000

Essay codes of the error-tagged subcorpus *BUCLE*_110,000							
essay code	**words**	**essay code**	**words**	**essay code**	**words**	**essay code**	**words**
BGSU1001	500	BGSU1046	1153	BGSU1093	365	BGSU1227	301
BGSU1002	502	BGSU1047	938	BGSU1094	330	BGSU1228	535
BGSU1003	779	BGSU1048	1020	BGSU1095	474	BGSU1229	452
BGSU1004	522	BGSU1049	460	BGSU1096	386	BGSU1230	422
BGSU1005	580	BGSU1050	464	BGSU1097	443	BGSU1231	385
BGSU1006	577	BGSU1051	829	BGSU1098	399	BGSU1232	417
BGSU1007	580	BGSU1052	369	BGSU1099	927	BGSU1233	911
BGSU1008	525	BGSU1053	554	BGSU1100	767	BGSU1234	512
BGSU1009	373	BGSU1054	470	BGSU1101	458	BGSU1235	692
BGSU1010	325	BGSU1055	990	BGSU1102	825	BGSU1236	427
BGSU1011	556	BGSU1056	1036	BGSU1103	1015	BGSU1237	441
BGSU1012	522	BGSU1057	594	BGSU1104	474	BGSU1238	351
BGSU1013	634	BGSU1058	592	BGSU1105	530	BGSU1239	415
BGSU1014	680	BGSU1059	850	BGSU1106	467	BGSU1240	405
BGSU1015	431	BGSU1060	1037	BGSU1107	528	BGSU1241	503
BGSU1016	216	BGSU1061	1087	BGSU1108	556	BGSU1242	369
BGSU1017	400	BGSU1062	459	BGSU1109	506	BGSU1243	591
BGSU1018	537	BGSU1063	382	BGSU1110	444	BGSU1244	674
BGSU1019	370	BGSU1064	1003	BGSU1111	364	BGSU1245	652
BGSU1020	349	BGSU1065	702	BGSU1112	351	BGSU1246	381
BGSU1021	446	BGSU1066	508	BGSU1113	300	BGSU1247	355
BGSU1022	621	BGSU1067	598	BGSU1114	898	BGSU1248	386
BGSU1023	473	BGSU1068	908	BGSU1115	920	BGSU1249	321
BGSU1024	514	BGSU1069	1012	BGSU1116	846	BGSU1250	457
BGSU1025	1030	BGSU1070	1030	BGSU1117	1169	BGSU1252	421
BGSU1026	437	BGSU1071	743	BGSU1118	452	BGSU1253	399
BGSU1027	346	BGSU1072	929	BGSU1119	550	BGSU1254	291
BGSU1028	439	BGSU1073	538	BGSU1120	1193	BGSU1255	704
BGSU1029	380	BGSU1074	700	BGSU1121	1002	BGSU1256	567
BGSU1030	618	BGSU1075	218	BGSU1122	430	BGSU1257	433
BGSU1031	544	BGSU1076	528	BGSU1128	279	BGSU1258	385
BGSU1032	249	BGSU1079	344	BGSU1135	405	BGSU1259	445
BGSU1033	253	BGSU1080	795	BGSU1136	557	BGSU1260	324
BGSU1034	534	BGSU1081	587	BGSU1137	1205	BGSU1261	606
BGSU1035	678	BGSU1082	533	BGSU1146	666	BGSU1262	460
BGSU1036	652	BGSU1083	609	BGSU1167	1030	BGSU1263	481
BGSU1037	512	BGSU1084	370	BGSU1207	1097	BGSU1264	355
BGSU1038	1204	BGSU1085	497	BGSU1218	1048	BGSU1265	401
BGSU1039	519	BGSU1086	523	BGSU1220	1043	BGSU1266	458
BGSU1040	461	BGSU1087	549	BGSU1221	1032	BGSU1267	637
BGSU1041	407	BGSU1088	475	BGSU1222	1126	BGSU1269	558
BGSU1042	670	BGSU1089	486	BGSU1223	948	BGSU1270	549
BGSU1043	2381	BGSU1090	352	BGSU1224	1037	BGSU1271	1021
BGSU1044	615	BGSU1091	389	BGSU1225	1504	BGSU1272	1061
BGSU1045	688	BGSU1092	442	BGSU1226	1292	BGSU1273	922
						BGSU1274	1312
Total number of essays: 181							
Total number of words: 112,064							

essay code	words	essay code	words	essay code	words	essay code	words	essay code	words
Essay codes of the error-tagged subcorpus _GICLE_110,000_								
DNNI5008	1190	GEAU1040	276	GEAU1090	258	GEAU2041	284	GEAU4004	399
FRUC1059	467	GEAU1041	336	GEAU1091	209	GEAU2042	284	GEAU4005	180
FRUL1002	705	GEAU1042	262	GEAU1092	443	GEAU2043	223	GEAU4006	487
FRUL1004	654	GEAU1043	283	GEAU1094	238	GEAU2044	279	GEAU4007	518
FRUL2001	623	GEAU1044	516	GEAU1095	374	GEAU2045	283	GEAU4008	421
FRUL2005	684	GEAU1045	484	GEAU1096	461	GEAU2046	281	GEAU4009	792
FRUL2007	520	GEAU1046	649	GEAU1097	571	GEAU2047	303	GEAU4010	1117
FRUL2008	541	GEAU1047	527	GEAU1098	446	GEAU2048	188	GEBA1011	408
FRUL2012	754	GEAU1048	195	GEAU1099	599	GEAU2049	265	GEBA1012	365
FRUL2013	470	GEAU1049	434	GEAU1100	689	GEAU3001	652	GEBA1013	378
FRUL2018	533	GEAU1050	339	GEAU1101	493	GEAU3002	895	GEBA1021	374
GEAU1001	461	GEAU1051	328	GEAU1102	299	GEAU3003	734	GEBA1029	765
GEAU1002	216	GEAU1052	400	GEAU1103	528	GEAU3004	483	GEBA1030	602
GEAU1003	240	GEAU1053	362	GEAU1104	317	GEAU3005	784	GEBA1031	997
GEAU1004	389	GEAU1054	236	GEAU1105	455	GEAU3006	797	GEBA1035	674
GEAU1005	364	GEAU1055	389	GEAU1106	558	GEAU3007	697	GEBA1039	449
GEAU1006	267	GEAU1056	339	GEAU1107	450	GEAU3008	456	GEBA1040	566
GEAU1007	451	GEAU1057	268	GEAU1108	433	GEAU3009	712	GEBA1041	422
GEAU1008	285	GEAU1058	237	GEAU2001	385	GEAU3010	875	GEBA1044	397
GEAU1010	254	GEAU1059	309	GEAU2002	190	GEAU3011	906	GEBA1045	492
GEAU1011	251	GEAU1060	413	GEAU2003	283	GEAU3012	810	GEBA1046	458
GEAU1012	386	GEAU1061	410	GEAU2004	240	GEAU3013	991	GEBA1047	488
GEAU1013	219	GEAU1062	364	GEAU2005	394	GEAU3014	886	GEDR1001	541
GEAU1014	252	GEAU1063	368	GEAU2006	297	GEAU3015	608	GEDR1002	999
GEAU1015	305	GEAU1064	281	GEAU2007	249	GEAU3016	718	GEDR1003	546
GEAU1016	281	GEAU1065	353	GEAU2008	343	GEAU3017	1209	GEDR1004	579
GEAU1017	356	GEAU1066	565	GEAU2009	333	GEAU3018	989	GEDR1005	521
GEAU1018	292	GEAU1067	374	GEAU2010	251	GEAU3019	718	GEDR1006	613
GEAU1019	267	GEAU1069	405	GEAU2011	388	GEAU3020	328	GEDR1007	716
GEAU1020	235	GEAU1070	262	GEAU2012	181	GEAU3021	1048	GEDR1008	632
GEAU1022	780	GEAU1071	391	GEAU2013	256	GEAU3022	449	GEDR1009	711
GEAU1023	625	GEAU1072	348	GEAU2014	299	GEAU3023	769	GEDR1010	835
GEAU1024	403	GEAU1073	435	GEAU2015	242	GEAU3025	879	GEDR1011	643
GEAU1025	596	GEAU1074	501	GEAU2016	287	GEAU3026	1018	GEDR1012	430
GEAU1026	625	GEAU1075	345	GEAU2017	237	GEAU3027	421	GEDR1013	871
GEAU1027	638	GEAU1076	367	GEAU2020	303	GEAU3028	696	GEDR1014	444
GEAU1028	458	GEAU1077	193	GEAU2021	257	GEAU3033	740	GEDR1015	528
GEAU1029	370	GEAU1078	469	GEAU2022	157	GEAU3040	815	GEDR1016	612
GEAU1030	327	GEAU1079	384	GEAU2024	221	GEAU3049	628	GEDR1017	633
GEAU1031	190	GEAU1080	268	GEAU2026	380	GEAU3050	959	GEDR1020	554
GEAU1032	350	GEAU1081	335	GEAU2030	198	GEAU3054	965	GEDR1021	589
GEAU1033	330	GEAU1082	288	GEAU2032	236	GEAU3057	530	GEDR1022	524
GEAU1034	308	GEAU1083	273	GEAU2035	279	GEAU3059	462	GEDR1023	638
GEAU1035	343	GEAU1084	293	GEAU2036	234	GEAU3062	604	GEDR1024	562
GEAU1036	393	GEAU1085	303	GEAU2037	186	GEAU3064	904	GESA4005	445
GEAU1037	209	GEAU1086	269	GEAU2038	172	GEAU3065	672	GESA5030	333
GEAU1038	305	GEAU1087	411	GEAU2039	169	GEAU3066	574	GESA5031	725
GEAU1039	256	GEAU1088	219	GEAU2040	261	GEAU3067	699	SWUL8005	561
		GEAU1089	316			GEAU3068	707		

Total number of essays: 241
Total number of words: 113,230

A 1.6 Aspect Errors and Corrections in BUCLE_110,000

Correction	Student Version	Target Version	Student Version
(VT: would have contributed)	has contributed something to it. But if the case i	would	has
(VT: had)	ever lived on this planet had used modern technology to write poems,	had lived	lived
(VT: had been a long time since she swapped, OR She had swapped the real world	was long ago	had been	was
(VT: had not been for...)	wasn't for this genious' imagination. He would have j	had not been for	was not for
(VT: have written)	once wrote cannot help you in a critical situation in your j	have written	wrote
(VT: have they taken shape, namely, become materialised)	did they actually take	have taken shape	did take shape
(VT: have taken)	took place. Whenever progress is measured, there is a driving	have taken	took
(VT: have signed)	signed it, consequently taking the responsibility to make sur	have signed	signed
(VT: have shown)	show how inefficient and impracticable that concept was. The q	have shown	show
(VT: have occurred)	occurred, all the innovations that (VT: have taken) took pl	have occurred	occurred
(VT: have noticed)	notice a praiseworthy tendency in our University teachers to	have noticed	notice
(VT: have not been)	are not the most suitable words for the description of what	have not been	are not
(VT: have made)	made, are making and will make in these areas is due to those w	have made	made
(VT: have made)	made their future our reality. But those who created our world	have made	made
(VT: have made it possible – though 'make' is also ok, just not as ok :)	make i	have made	make
(VT: have kept)	keep so jealously. Nowadays some things have changed. We study,	have kept	keep
(VT: have invented)	invented help us. They are so delightful, so immaculate, th	have invented	invented
(VT: have found that)	found that there are two kinds of imagination. We may cal	have found	found
(VT: have expressed)	expressed my opinion about the importance of the universit	have expressed	expressed
(VT: have developed)	'develop your imagination'?" It sounded impractical, an ec	have developed	develop
(VT: have destroyed)	destroyed everything we possessed. So following the logic	have destroyed	destroyed
(VT: have come to find it, or: have started to find it)	find it extremely profi	have come to find	find
(VT: have come; or: come)	came to money again although trying to escape it, we	have come	came
(VT: have begun)	began to realize that something (VT: has been lost) was lost i	have begun	began
(VT: have been)	were very few genius scientists and inventors who (VT: have acc	have been	were
(VT: have been)	were blissfully happy in their lives. No matter who one lives h	have been	were
(VT: have been)	are fulfulled. But as far as the curriculum is concerned there	have been	are

(VT: have been)	were made that it is not a surprise that we are often said to l	have been	were
(VT: have been sent)	were sent to the Space, vaccines against lethal diseases (	have been	were
(VT: have been offered; or, simply: '...in which they live...')	are offered to	have been	are
(VT: have been built)	were built. The machines we (VT: have invented) invented	have been	were
(VT: have been)	are more or less the same since Ancient Greece. On the other ha	have been	are
(VT: have been)	were made, on the political level, for example, communism was s	have been	were
(VT: have been discovered)	were discovered, bridges (VT: have been built) were	have been	were
(VT: have been)	are so improved that they can operate even without the interfer	have been	are
(VT: have already been travelling)	used to travel and (VT: will continue to tra	have been	used to
(VT: have already been forgotten)	are already forgotten. Today barely anyone kn	have been	are
(VT: have accumulated)	accumulated there will hardly ever be useful to them in	have accumulated	accumulated
(VT: have accomplished)	accomplished scientific breakthroughs and contributed a	have accomplished	accomplished
(VT: has remained)	remains unchanged since a certain stage of the evolution. Th	has remained	remains
(VT: has proved)	proved and the present is still proving that the sensitive non	has proved	proved
(VT: has only read)	had only read and never before operated in his life I would	has read	had read
(VT: has now been changed)	is now changed due to the development of the science	has been changed	is changed
(VT: has not changed)	does not change in its philosophical approach to life. It	has not changed	does not change
(VT: has lived)	lives, there have always existed different types of people, wit	has lived	lives
(VT: has issued)	issues a warning that unless something is done, the human race	has issued	issues
(VT: has)	have never before thought about; industrialisation, a product of the	has thought	have thought
(VT: has happened)	happenned around and so man made up his mind to found "schoo	has happened	happened
(VT: has)	had never made translations before; or a lawyer wouldn't (VT: will no	has made	had made
(VT: has)	had never, at least, assisted in such an activity// Nobody learns how	has assisted	had assisted
(VT: has fused)	fuses to a great extent with the dry science. Let me take for a	has fused	fuses
(VT: has ceased)	ceased to strike us as unusual and innovative. As example is H	has ceased	ceased
(VT: has been a technologist)	is therefore a technologist from the beginning, a	has been	is
(VT: has been)	were openly accused of inciting the conflicts. What matters is t	has been	were
(VT: has been completely forgotten about)	is completly forgot about since all	has been	is
(VT: has been lost)	was lost in our education. Didn't it turn to Medieval theor	has been	was

(VT: has become)	becomes for them an age of Romaniticism. A late one. Despite	has become	becomes
(VT: has become)	became a cause for frustration and even fear. Consequently, th	has become	became
(VT: have been made)	been made, if man 10. /had not used/ his imagination? It i	have been	have' omitted
(VT: The time has come that...)	It is time, that our society is dominated by in	has come	it is time
(VT: who lived)	who have lived long ago. The ancient, for instance , did not be	lived	have lived
(VT: who discovered)	who had discovered the laws of mechanics and gravitation.	discovered	had discovered
(VT: were)	have been a vision in somebody's dreams , provoked by the necessity	were	have been
(VT: were)	have been among the basic tools used for this fulfillment. Here I wo	were	have been
(VT: were)	had been once (not long ago) just projects taking shape in the minds	were	had been
(VT: were even able)	have even been able to condense it into a small box, calle	were	have been
(VT: went)	have been to, (VT: helped) has helped me travel in the world of the	went	have been
(VT: went)	had gone out in the forest to chop wood (VT: have already been forgo	went	had gone
(VT: we met)	we've met the other day. We manage to do this with the help of out	met	have met
(VT: washed)	had washed the clothes in the river with their hands bleeding and	washed	had washed
(VT: was; though, in some senses 'has been' could be ok...)	has been brought up	was	has been
(VT: was then impossible; or: what used to be impossible)	had then been impossi	was	had been
(VT: was later materialised)	has later been materialized into a stepping-stone	was	has been
(VT: was)	has been more need of workers now the reverse situation can be observ	was	has been
(VT: was)	has been obliged to learn during his studies. The reverse is equally	was	has been
(VT: was)	has been much more than it is now needed to perform the same task. Th	was	has been
(VT: was)	has been severed. Our grandparents' generation saw the end of a thous	was	has been
(VT: was)	has been only a vision in one's mind (VT: was later materialised) has	was	has been
(VT: was)	has been the major concern of governments, religions, political parti	was	has been

(VT: was)	has been the force necessary for development. Nowadays, in the techno	was	has been
(VT: visualised)	have visualized the innovations of their dreams: how they woul	visualised	have visualised
(VT: understood)	have understood earlier that theory without practice counts fo	understood	have understood
(VT: tried)	have tried to write (VT: helped) have helped me find my own truth -	tried	have tried
(VT: took /up/)	has taken the space of a whole room; while nowadays we can hold	took	has taken
(VT: took place)	have taken place a slight degree. Over 80 per cent of manual w	took	has taken
(VT: took)	had taken (VT: have shown) show how inefficient and impracticable th	took	had taken
(VT: tinted)	have tinted my grey daily round with a " verse rainbow ". Yes, I a	tinted	have tinted
(VT: threw)	thrown it away or even quite (VT: forgot) forgotten about it? Maybe	threw	have thrown
(VT: studied)	have studied the same things - what the professors told them in t	studied	have studied
(VT: studied)	had studied for. With all that in mind you soon end up asking you	studied	had studied
(VT: stepped)	had stepped on it. Still, the stars and the Moon remain the dista	stepped	had stepped
(VT: spent)	have spent a few years in the US. Of course this is a little far -f	spent	have spent
(VT: saw)	have seen a cartoon recently, but I don't know its name. I do remembe	saw	have seen
(VT: sacrificed)	have willingly sacrificed the power of magic inside them and (	sacrificed	have sacrificed
(VT: really had)	has really had this imagination. Although he could not see the	had	has had
(VT: reached)	had reached a point when knowledge was no longer mythological but	reached	had reached
(VT: passed)	have passed till man discovered simple truths as the fact that the	passed	have passed
(VT: painted)	has painted its most vivid pictures, its most beautiful images. T	painted	has painted
(VT: met)	have met their boyfriend/girlfriend (wife/husband) through the Intern	met	have met
(VT: loaded)	has loaded the memory with this lore. Imagination is again the ori	loaded	has loaded
(VT: I only found ...)	I've only found the answer a few days ago. Simply as tha	found	have found
(VT: helped)	have helped me find my own truth - that is,"creative writing has b	helped	have helped
(VT: helped)	has helped me travel in the world of the subconscious. I (VT: expe	helped	has helped

(VT: helped)	has helped development in the field of engineering. Not that parad	helped	has helped
(VT: had to take)	take. Computers are becoming more and more important in our l	had to take	take
(VT: had to do)	had had to do only to procure the heat of the rumbling stove or	had to do	had had to do
(VT: graduated)	have graduated a couple of years ago. The most natural thing fo	graduated	have graduated
(VT: graduated)	has graduated from, even one's nationality and religion. And th	graduated	has graduated
(VT: forgot)	forgotten about it? Maybe the technical luxury in which they (VT:	forgot	forgotten
(VT: first came alive)	has first become alive in man's dreams. What (VT: was) h	came	has come
(VT: experienced)	have experienced poetry as the catharsis of all negative emot	experienced	have experienced
(VT: ever was)	has ever been. Last century was one of inventions and rapid prog	was	has been
(VT: ever learnt)	had ever learned literature with a textbook in his hand. This	learnt	had learnt
(VT: didn't / did not)	hadn't wish to industrialize and modernize? I mean - how	did not wish	had not wish
(VT: did not take or had not taken)	had not taken into account. As a whole, I w	did not take	had not taken
(VT: did)	have done there, all the poems I (VT: tried) have tried to write (VT:	did	have done
(VT: we are subject to; or: we are being subject to)	we've been subject to, at	are	have been
(VT: often ask)	have often asked myself " what does " dreaming mean to me? " It	ask	have asked
(VT: is)	has been no place left for dreaming and imagination is a result from t	is no place	has been no place
(VT: is)	has been ridiculed now and again. Thus a strange phenomenon is observe	is ridiculed	has been ridiculed
(VT: go)	have gone to such troubles and efforts trying to get into an universit	go	have gone
(VT: exceeds)	has exceeded the technical power of its time , its products can l	exceeds	has exceeded
(VT: abhorred)	have abhorred during your first year) in order to solve a minor	abhor	have abhorred
(VT: will turn)	are turning into machines - rational, pragmatic, cold. Let's no	will	are turning
(VT: he had been working)	worked since 1665 and be present at a meeting which h	had been working	worked
(VT: always wondered)	been always wandering about the world, our country and th	have wondered	have been wondering
(VT: have been creating)	create paintings, songs and other objects of art that	have been creating	create
(VT: have been robbing)	robbed their own country and its people, contributing t	have been robbing	have robbed
(VT: been developing in terms of...)	developed and are constantly developing in	have been developing	have developed
(VT: have dealt)	been dealing with the question: 'Are men equal?' and if so why	have dealt	have been dealing
(VT: has existed)	has been existing for quite a long time. But how did it all b	has existed	has been existing

(VT: searching)	search inspiration and ideas. We dodge the hits of life and we	are searching	search
(VT: is turning)	turns out to be quite a controversial one in terms of historic	is turning	turns
(VT: is travelling)	travels through the desert in South Africa. Suddenly it bre	is travelling	turns
(VT: is taking or has taken)	takes big steps forward, as well: there is a whole	is taking	takes
(VT: is ruining)	ruins my fragile inner world. I have no free time to devote to	is ruining	ruins
(VT: is he referring)	does he refer to the world and not to Bulgaria. The point	is he referring	does he refer
(VT: is enjoying)	enjoys its zenith, more and more people are worshipping their	is enjoying	enjoys
(VT: is coming)	has come to its close we seem more and more infected with the e	is coming	has come
(VT: is becoming)	becomes more and more widely-used by millions of people. Of c	is becoming	becomes
(VT: being develop)	developed; research provides with new and interesting insig	are being developed	are developed
(VT: are taking)	take interest in art. " But has technology really replaced art	are taking	take
(VT: are studying)	study English philology, had to go through a course in infor	are studying	study
(VT: are studying for)	study for, in the first place. In Bulgaria, we know that	are studying	study
(VT: are starting)	start to see a light at the end of the tunnel.	are starting	start
(VT: are seeing)	see is really happening or it's just a dream? Now that you're	are seeing	see
(VT: are replacing, OR: have replaced)	replace humans in almost all branches of	are replacing	replace
(VT: are pushing)	push their planet too hard and are about to blow it up. There	are pushing	push
(VT: are preparing)	prepare to work in various spheres of our multidimensional	are preparing	prepare
(VT: are growing)	grow bigger and bigger. This statement reminds me of another	are growing	grow
(VT: are getting)	get more free. I (VT: have noticed) notice a praiseworthy ten	are getting	get
(VT: are getting, or: have gotten)	get more and more estranged. We have almost	are getting	get
(VT: are experiencing)	experience now, three centuries later. Historians claim	are experiencing	experience
(VT: are either running away)	either run away from something, or (VT: searching	are running	run
(VT: are dreaming of)	dream how to invent a new kind of light bulb that will no	are dreaming	dream
(VT: are becoming / getting / ending up / etc)	are more and more remote to drea	are becoming	are more and more
(VT: are answering)	answer it negatively. Those people educate their children t	are answering	answer
(VT: am teaching)	teach two friends of mine English and I can say without exagg	am teaching	teach
(VT: am expressing, or: have been expressing)	expressed my personal belief that	am expressing	expressed
(VT: are going to talk about, OR: if we were to talk about)	should talk about c	are going to talk	should talk
(VT: dreamt, OR: used to dream)	were dreaming when we were kids? I think not. I	dreamt	were dreaming
(VT: what you specialise in)	are you specializing in , because the way things a	specialize	are specializing

(VT: think about)	are thinking about its replacement. "Mammon" as Carlyle names	think	are thinking
(VT: study, or: have studied))	are studying at university or not. One of the gr	study	are studying
(VT: stated)	being continuously officially stated and re-stated often the talk	is stated	is being stated
(VT: shake)	are shaking nervously, before the results are announces. How much t	shake	are shaking
(VT: organised)	being organised, we will soon notice a certain formula shared b	is organised	is being organised
(VT: live)	are living and very often by the idea that blacks are inferior. Desp	live	are living
(VT: live)	are living they become more and more different from each other, uneq	live	are living
(VT: is, or: has been)	being tasted, there is nothing that can top or deviate t	is tasted	is being tasted
(VT: is encouraged)	is being encouraged by certain modern trends in literature	is encouraged	is being encouraged
(VT: daydream)	am Daydreaming and still I believe this is what makes me a human	daydream	am daydreaming
(VT: comes, or: can come)	is coming our way in the form of the project for a di	come	is coming
(VT: attend)	attending lectures and seminars every day. When I was a child I us	attend	are attending
(VT: ascribed)	being ascribed financial value. Contribution to society, however	is ascribed	is being ascribed
(VT: are examined)	are being examined and cured, even correct things in their o	are examined	are being examined
(VT: are being printed)	are printed and published here in Bulgaria, we can easi	are printed	are being printed
(VT: already perceives)	is already perceiving the things around him, and as lon	perceives	is perceiving

Aspect Errors and Corrections in GICLE_110,000

Correction	Student Version	Target Version	Student Version
(VT: can take place)	has taken place. They want a change here and now. They forg	can take	has taken
(VT: would make)	have made in this case (VT: was) is that when I (VT: listen)	would make	have made
(VT: had been mapped out)	was mapped out considerable time before it was passed	had been mapped out	was mapped
(VT: had been)	was prepared the evening before? I used to think that it was a go	had been	was prepared
(VT: had been)	was three hours before. When I finally arrived it was four o'cloc	had been	was
(VT: had been)	were forecasted, there came more than one thousand visiters from	had been	were
(VT: had been)	was one. Around September I realized that I hadn't succeeded in	had been	was
(VT: had complained)	complained about the nuisance cars made. They didn't dare t	had complained	complained
(VT: had had)	had in the local beer-garden of Mutzenwinkel, a village with less	had had	had
(VT: had joined)	joined me on the occasion of a march `against a new Nazi-age in	had joined	joined
(VT: had made, OR: was making)	made, that his trainer was very proud of him and	had made, OR: was making	made
(VT: had returned)	returned at all, "danced" around me and barked as loud as he	had returned	returned
(VT: had started)	started spreading rumours about me. It came to the point where	had started	started
(VT: had started)	started out fine. It was springtime when I moved into my litt	had started	started
(VT: had wanted to)	wanted to spend some days out in the green - he yells at her	had wanted to	wanted to
(VT: had not)	wouldn't have gained our attention, we never would have recognized	had not	would not have
(VT: had been sitting)	were sitting for just a few minutes as something happened	had been sitting	were sitting
(VT: had been watching)	was watching a crime story that evening and now Angela w	had been watching	was watching
(VT: had been including)	includes purely fictional reports. Since these anti-li	had been including	includes
(VT: done)	did more harm than good to mankind. That is the same with genetic eng	done	did
(VT: has agreed to call, OR: calls typical)	agreed to call typical housewives' w	has agreed to call, OR: calls typical	agreed
(VT: has already been)	was already the third day on which I displayed my womanly	has already been	was already
(VT: has already given)	alreay gave birth to four children, that she looks after	has already given	already gave birth
(VT: has always believed)	believed in the truthfulness of visual data. The audi	has always believed	believed
(VT: has become)	became clearly visible in the last two decades and has not reac	has become	became
(VT: has been proved)	was proved by independent tests several times. The watcher	has been proved	was proved
(VT: has been) abl	was locked all day, admire the Christmas tree with its predict	has been abl	was locked
(VT: has been)	is proven that musicians themselves sometimes commit awful crimes	has been	is proven

(VT: has been proved)	was proved by independent tests several times. The watcher	has been proved	was proved
(VT: has been abolished)	was abolished in most western countries. Legislators ar	has been abolished	was abolished
(VT: has been known)	is known to the public at least since the asylum debates at	has been known	is known
(VT: has been)	is allowed to sell gene-food, for instance manipulated tomatoes,	has been	is allowed
(VT: has been)	was a failure. Nobody, neither my mother nor the men I loved, (V	has been	was
(VT: has broken out)	broke out like a cleaning-mania, a movement that is whizzin	has broken out	broke
(VT: has brought)	brought the perception of women in nowadays society into the c	has brought	brought
(VT: has changed)	changed our lives fundamentally. Today everyone just has got o	has changed	changed
(VT: has come to play)	became to play the centre part in every's life and that i	has come to play	became
(VT: has declared)	declared "We have to stop traffic in the city center. Due to	has declared	declared
(VT: has decreased)	decreased. So, if you don't want to be the next victim of a	has decreased	decreased
(VT: has either served or not served)	either served or did not serve and (VT: do	has either served or not served	served
(VT: has entered)	entered into my dream. She (VT: can) could hear her husband st	has entered	entered
(VT: has even become)	even became worse. And if people don't starve they're kill	has even become	became
(VT: has had)	had a heart attack. Furthermore the telephone offers the possibili	has had	had a heart attack
(VT: has just opened)	just opened the door holding a long and shiny knife in his	has just opened	opened
(VT: has learnt how to communicate with)	learned communicating with other people	has learnt how to communicate with	learned
(VT: has made)	made some people aware that there are too many cars in the world.	has made	made
(VT: has made)	made a deeper understanding of women's situation possible. Even t	has made	made
(VT: has really changed)	really changed our lives and we can be grateful about i	has really changed	really changed
(VT: has stated)	stated that there is almost no sense in this work because when	has stated	stated
(VT: has succeeded)	succeeded in achieving a change, slight it may be. I am a ho	has succeeded	succeeded
(VT: has thought)	thinks of abandoning the Olympic Games so far so we will have	has thought	thinks
(VT: have already had)	had allready three smog alarms this year and of nothing c	have already had	had already three
(VT: have already got used to)	got used to it and do not care anymore. I someti	have already got used to	got used to
(VT: have also had)	also had such an experience. It was in the summer of 1990. I	have also had	also had an experience
(VT: have been replaced)	are replaced by buckets filled with gloriously colourfu	have been replaced	are replaced
(VT: have been launched)	were launched some unbelievably cruel attacks on foreig	have been launched	were launched
(VT: have been built)	were built up in America where violators with psychopathic	have been built	were built

(VT: have been)	were diminished or eliminated successfully throughout the centur	have been	were diminished
(VT: have been laid)	are laid to rest. Second the noise of the cars is a danger	have been laid	are laid to rest
(VT: have been organised)	were organized allover Germany, not only in huge citie	have been organised	were organized
(VT: have been told; were told; or are told)	were told to collect materials sepa	have been told; were told; or are told	were told
(VT: have been)	are destroyed, and he has to show all his nuclear and chemical w	have been	are destroyed
(VT: have been built)	are build new hospitals in Kreischa. Thousands of jobs for	have been built	are build
(VT: have been destroyed)	are destroyed according to the last poll. Rare plants,	have been destroyed	are destroyed
(VT: have begun, OR: are beginning)	begin to feed us with such stupid programs c	have begun, OR: are beginning	begin to feed
(VT: have bought)	buy the new product regardless of the quality of this new prod	have bought	buy the new
(VT: have come to feel)	feel that cycling not only keeps me in good exercise but	have come to feel	feel that
(VT: have completely given up)	completely gave up going on package holidays. Eve	have completely given up	gave up
(VT: have)	decided to help in Somalia - despite the looming danger of assassinat	have decided	decided
(VT: have described)	described several times. In most cases I was successful in	have described	described
(VT: have developed)	developed, (VT: has been) was a failure. Nobody, neither m	have developed	developed
(VT: have died)	died of an overdose. These must simply be one preventive measure	have died	died
(VT: have enjoyed)	enjoyed better education are able to pick up the information	have enjoyed	enjoyed
(VT: have hacked and slashed)	hacked and slashed our bloody way through history.	have hacked and slashed	hacked
(VT: have had)	had pleasure are treated like treasury but never mind - you'll te	have had	had pleasure
(VT: have happened)	happened the last 24 hours. "Do you think that Peter wants t	have happened	happened
(VT: have just said)	just said that I did not want to take my bike today but my	have just said	just said
(VT: have lost)	lost so many traditions, why should we also break with the tradi	have lost	lost
(VT: have made)	made, every method I (VT: have developed) developed, (VT: has	have made	made
(VT: have made)	made the same experience and therefore understand him or they th	have made	made
(VT: have made – though to be correct, it should be: have had)	made, one sticks	have made – though to be correct, it should be: have had	made
(VT: have managed)	managed to force society to reconsider its values and especia	have managed	managed
(VT: have never learnt)	learnt how to deal with the mass media, with the amount	have never learnt	learnt
(VT: have not had)	didn't have it yet. It goes without saying that it is impossi	have not had	did not have it
(VT: have obviously managed)	obviously managed to bring about a revolution in th	have obviously managed	managed

(VT: have often had)	had discussions with my mother concerning whether I should	have often had	had discussions
(VT: have predicted)	predict that human life can be radically prolonged. Would y	have predicted	predict
(VT: have realised)	realize that the success of these experiments have many cons	have realised	realize
(VT: have rented, OR: rent)	rented a terraced house in the countryside we live v	have rented, OR: rent	rented
(VT: have substituted)	substituted those which were ill or were not functioning.	have substituted	substituted
(VT: have succeeded)	succeeded in doing so and I'm really happy about it. With	have succeeded	succeeded
(VT: spend, OR: have spent)	spent for TV commercials due to consumers that are a	spend, OR: have spent	spent
(VT: that is has crept)	who creeps unconsciously into our life. It represents ab	that is has crept	creeps
(VT: have been discussing)	discuss about this problem in the city hall. Let us h	have been discussing	discuss
(VT: have been trying to)	try to find out whether animals (VT: were) are envolve	have been trying to	try to find
(VT: always used to make fun of)	had always been making fun of her old-fashioned	used to	had been
(VT: cost)	had cost and they even (VT: knew) know the day of my brothers's marri	cost	had cost
(VT: did)	have done a wonderful job. Exactly one year later when I came to see t	did	have done
(VT: didn't)	haven't you? And there is this young man with his long fair hair b	did	have
(VT: forgot)	had forgotten that they (VT: had wanted to) wanted to spend some da	forgot	had forgotten
(VT: forgot; OR: did forget)	had forgotten ... Liza didn't want to think about	forgot; OR: did forget	had forgotten
(VT: graduated)	had graduated in communication studies, it was impossible for he	graduated	had graduated
(VT: happened)	had happened 2000 years ago and may (VT: find) have found (apart	happened	had happened
(VT: has been; OR: was – it depends on when this essay was written, when the boo	had been a bestseller	was	had been
(VT: I learnt)	I've learned a lot in Ireland. Especially, (VT: I learnt) I've l	I learnt	have learned
(VT: I learnt)	I've learned to dress in the way I liked without thinking of othe	I learnt	have learned
(VT: identified)	has identified her mirror image as a picture of herself. This a	identified	has identified
(VT: knew)	you've known it anyway, (VT: didn't) haven't you? And there is this	knew	have known
(VT: lasted)	had lasted for almost half a century. However, those who indulged	lasted	had lasted
(VT: left)	had left her native beloved country to find adequate work abroad. Alt	left	had left
(VT: made)	have made us forget the fear of a nuclear war that had been lurking i	made	have made
(VT: met)	have met there would be willing and would be glad to make use of	met	have met

	their		
(VT: noticed)	had noticed on numerous occasions that there was hardly a town in	noticed	had noticed
(VT: organised)	had organized an open-air-concert, where they performed ethnic d	organised	had organized
(VT: put)	has put an end to the division of the world that (VT: lasted) had last	put	has put
(VT: sent)	has sent the kids to bed admonishing them to keep quiet as Mummy was	sent	has sent
(VT: showed)	has showed that highly developed animals have a awareness of themse	showed	has showed
(VT: sold)	has sold his car and now he (VT: was using) uses his bicycle or publi	sold	has sold
(VT: they chose)	chosen Ronald MacDonald their favourite celebrity in 1992; Mich	they chose	have chosen
(VT: told)	had told me with a smirk. So I had passed a whole evening chatting wi	told	had told
(VT: told)	has cheerfully told me the experiences he (VT: had had) had in the l	told	has told
(VT: told)	have told a lot of facts about experiences of scientists who (VT: had	told	have told
(VT: tolf)	had told them a fairy-tale before - may be the one of the little boy	told	had told
(VT: undertook)	have undertaken to anyone else. It (VT: turned out) was going t	undertook	have undertaken
(VT: used to be)	had used to be an interesting and extraordinary thing but now p	used to be	had used to be
(VT: was followed)	had been followed by many others, and after a time I began to	was followed	had been followed
(VT: was)	had been! Only town dwellers like me who have never been to the countr	was	had been
(VT: was)	had been an enormous success, and I'm sure that it was a great cultura	was	had been
(VT: was)	had been his death sentence. But what will happen to the RAF terrorist	was	had been
(VT: was)	had been in the time of my childhood. Whenever I (VT: come) came home	was	had been
(VT: was)	has been a big sensation. There was and are a lot of praise and enthus	was	has been
(VT: was)	has not been that long ago that I left school and my reason for choosi	was	has been
(VT: was)	has also been one of the grey little boxes that broke the news about m	was	has been
(VT: was)	has been a generation ago. After the two wars there was an upward move	was	has been
(VT: was)	have been born this way. Yes. I think that's why I can't remember when	was	have been
(VT: was in / joined)	have joined the army for 12 months, I have experience enou	joined	have joined

(VT: was known)	has been known as the bad old days of imperialism, many European	was	has been
(VT: was only the phone)	has only been the phone on our floor in the students' h	was	has been
(VT: wasn't)	hadn't been too fine and heavy thunderstorms (VT: had been) were fo	wasn't	hadn't been
(VT: went to)	had been to the formal ball the mayor of our town had organized to	went	had been
(VT: Were you there yesterday)	Have you been yesterday on the report on the grow	were	have you been
(VT: accompany them; OR are accompanying them)	have accompanied them in order to	accompany them	have accompanied
(VT: allows)	has allowed us to cope with diseases like tuberculosis, pneumonia a	allows	has allowed
(VT: arrive)	have arrived the middle of the street, suddenly a car seems to come	arrive	have arrived
(VT: celebrate)	have celebrated Christmas in memory of what (VT: happened) had h	celebrate	have celebrated
(VT: eat)	had eaten them I (VT: don't / do not) didn't become thirsty at all. It	eat	had eaten
(VT: enjoy)	have enjoyed some didactics, but by no means enough compared to our	enjoy	have enjoyed
(VT: enjoyed)	has enjoyed this exotic behaviour and immediately took part in som	enjoyed	has enjoyed
(VT: find)	have found (apart from the presents under the Christmas tree) hope de	find	have found
(VT: is not by any means complete)	hasn't been by any means completed. Another	is not by any means complete	has not been
(VT: lands)	has landed on the floor for the first time, you will either keep an	lands	has landed
(VT: reads)	has read the example about my father, (VT: probably thinks) is proba	reads	has read
(VT: realise)	have realized that they don't really object to the traditional par	realise	have realized
(VT: had been including)	includes purely fictional reports. Since these anti-li	had been including	includes
(VT: had been flourishing)	had flourished not only in the United States, but had	had been flourishing	had flourished
(VT: enjoying)	enjoyed cycling out there in fresh air and in bright sunshine. An	enjoying	enjoyed
(VT: he was referring)	referred to Rock music in general. I must admit that when	he was referring	referred
(VT: was asking)	ask, and that means it (VT: would be) is far better to answer,	was asking	ask
(VT: was coming)	come to see me and when: "Oh, by the way, was that your new boy	was coming	come
(VT: was enjoying)	enjoyed this practicable and quick way of eating. The book w	was enjoying	enjoyed
(VT: was hurrying)	hurried through the city to get at least a few of the mountai	was hurrying	hurried
(VT: was leafing)	leafed through a weekly magazine recently, my eyes were magica	was leafing	leafed
(VT: was not referring)	did not refer to heavy metal or trash metal, which is in	was not referring	did not refer
(VT: was really looking forward)	looked really forward to my breakfast. The menu	was looking	looked

(VT: was she suffering)	does she suffer from a terrible malady? Dot, dot, dot.	was suffering	does suffer
(VT: was working)	worked in such a place in order to earn money in her spare tim	was working	worked
(VT: were being exploited)	were exploited and if they knew that they could chang	were being exploited	were exploited
(VT: were making)	made derogatory allusions to those who fuelled themselves with	were making	made
(VT: were saying)	said. But wasn't there still something else in this well-known	were saying	said
(VT: were trembling, or started to tremble)	trembled and I was as white as a she	were trembling	trembled
(VT: When he was wanting to get out)	As he wanted to get out a big dog approache	was wanting	wanted
(VT: have been witnessing)	have witnessed what was intended to be a mission of c	have been witnessing	have witnessed
(VT: have been discussing)	discuss about this problem in the city hall. Let us h	have been discussing	discuss
(VT: have been trying to)	try to find out whether animals (VT: were) are envolve	have been trying to	try to find
(VT: has done to the mountains so far; OR: has been doing to the mountains.)	has been doing to the mountains	has done to the mountains so far;	has been doing
(VT: am speaking)	speak about health resorts I come to the most important point,	am speaking	speak
(VT: are beginning)	begin to succeed in their strategy of signing bands and buil	are beginning	begin
(VT: are doing)	do something wrong when they copy the heroes' behaviour. There a	are doing	do
(VT: are earning)	earn less money than others. Why does it seem so easy for othe	are earning	earn
(VT: are eating)	eat. They want to have a sign on food which is manipulated. 90%	are eating	eat
(VT: are falling)	fall victim to the growing pollution. Frank Huber the leading	are falling	fall
(VT: are getting)	get another perspective and another shaping as members of the	are getting	get
(VT: are losing)	lose more and more the image of a simple, naive house-wife, bec	are losing	lose
(VT: are planning)	plan their career in the upper floor of big companies. They i	are planning	plan
(VT: are skating)	scate on very thin ice! Nobody really knows what the situation	are skating	scate
(VT: are still buzzing)	buzz around like workaholics-depending on their average	are buzzing	buzz
(VT: are taking)	take over full responsibility for theirselves, their job and th	are taking	take
(VT: are tending)	tend to adopt the capitalistic and democratic system of Wester	are tending	tend
(VT: are trying)	try to show their courage. But these things that you can see on	are trying	try
(VT: are trying or have been trying)	try to stop this damage of the environment	are trying	try
(VT: are waiting)	wait for the tram on a Saturday morning in order to go shoppin	are waiting	wait
(VT: can try; are trying; or try)	try to resolve the problem by not using or red	are trying	try

(VT: is becoming)	becomes increasingly important for international communication	is becoming	becomes
(VT: is building up)	is build up and you don't even dare to go to the toilet inb	is building up	is build up
(VT: is covering)	covers our country. This fact reduces the space of animals to	is covering	covers
(VT: is eating)	eats. And we like the minced meat enriched with E 205 and the so	is eating	eats
(VT: is facing)	faces one of its biggest problems. Car-parking and the rapidly i	is facing	faces
(VT: is increasing)	increases faster than they (VT: can) could renovate the plac	is increasing	increases
(VT: is nursing)	nurses crying babies to make them happy? Who has to coordinate	is nursing	nurses
(VT: is rushing)	rushes through the supermarket a quarter to six, to buy food fo	is rushing	rushes
(VT: is slowly taking over)	slowly takes over the former role of the church: the	is slowly taking over	takes over
(VT: is trying to)	sell life insurances by telephone? Will I have the chance to	is trying to	sell
(VT: is wishing for cars to be banned)	wishes that cars ought to be banned. The	is wishing for cars to be banned	wishes
(VT: sashaying over)	he sashays over the road, because he doesn't know that he i	sashaying over	sashays
(VT: broadcast)	were broadcasting the horryfying pictures from Hoyerswerda: mili	broadcast	were broadcasting
(VT: dealt)	was dealing with these huge forests in America which were all cut do	dealt	was dealing
(VT: developed)	was developing, and people were optimistically looking towards a	developed	was developing
(VT: flourished)	were flourishing, technology (VT: developed) was developing, a	flourished	were flourishing
(VT: oozed)	was oozing through the paper onto my palm. Feeling deeply disappoin	oozed	was oozing
(VT: sat)	was sitting on my knees. And I think they enjoyed my company as well b	sat	was sitting
(VT: started to exploit)	were evidently exploiting the accumulated fertility of	started to exploit	were exploiting
(VT: that still reminded)	that's still reminding him of his grand-mother. He (VT	that still reminded	is still reminding
(VT: watch)	were watching the Games certainly have bread and more than that: Als	watch	were watching
(VT: awaits)	is awaiting her - hot arguments between her parents, her bored elde	awaits	is awaiting
(VT: becomes)	is becoming clearer, I begin to understand the reason for my stran	becomes	is becoming
(VT: breathes or can breathe)	is breathing fresh air instead of exhaust fumes. H	breathes	is breathing
(VT: broadcast)	are broadcasting advertisments, simply by interrupting the film.	broadcast	are broadcasting
(VT: call)	are calling me three times a week although I don't want to chat with	call	are calling
(VT: comes)	is coming into my mind. How about buying my own telephone so I (VT:	comes	is coming

(VT: depend)	are depending on the money from the tourism. I remember one day, w	depend	are depending
(VT: develop)	are developing as fast as the crime does, there (VT: will there b	develop	are developing
(VT: do not just attract opponents)	are not just attracting opponents among men	do not attract	are not attracting
(VT: does not ring)	is not ringing. Or the other way round when I want to call m	does not ring	is not ringing
(VT: dream)	am dreaming about past times I see little villages surrounded by dar	dream	am dreaming
(VT: drinks)	is drinking, the children (VT: go) are going their own ways. Relig	drinks	is drinking
(VT: drives)	is driving everywhere with his car. This is only to feel oneself co	drives	is driving
(VT: fight over)	are fighting with a game boy. These mass medias and computer g	fight over	are fighting over
(VT: get)	are getting smaller and darker. There it is! Excitement, joy and a fee	get	are getting
(VT: go)	am going to the university by train. Opposite me there is a young man.	go	am going to the uni
(VT: go)	are going their own ways. Religion, traditions, conventions only serve	go	are going their own
(VT: go)	are going by train, car or aeroplane and (VT: travel) travelling all ar	go	are going by train
(VT: go)	are going there by car. "The car is man's	go	are going by car
(VT: have)	are actually having some relation to their relations - and I can't ex	have	are having
(VT: I go)	I'm going into the city very often now. I have even started doing som	I go	am going to the city
(VT: I lie)	I'm lying on my bed, a pot of tea and a plate of biscuits are next t	I lie	am lying
(VT: is served)	is being served everything from traditional german dishes with t	is served	is being served
(VT: is together)	is staying together and they are ready to go shopping. The car	is together	is staying together
(VT: leaps and jumps)	is leaping and jumping like a tennisball. At last, I alway	leaps and jumps	is leaping
(VT: lie)	are lying in man itself. Everybody has dark sides but most people try	lie	are lying
(VT: lies)	is lying on the ground, moving slowly. Maybe he is suffering from kni	lies	is lying
(VT: listen)	am listening to music, I am very interested in, I cannot concen	listen	am listening
(VT: makes)	is making its way through the town, transforming it completely. Brak	makes	is making
(VT: play)	are playing with horses, dogs and sheep. They call their favourite ba	play	are playing
(VT: praise)	are praising all kinds of vegetables and fruits, meat, fish and che	praise	are praising
(VT: probably thinks)	is probably thinking my father is crazy, or they (VT: have	probably thinks	is probably thinking
(VT: produce)	are producing by yourself. Wertstoffhof - that's also a password!	produce	are producing
(VT: react)	are reacting to the demand of raw material ranging from flower, seed	react	are reacting

(VT: rule)	are ruling the world. Man is his own worst enemy as Cicero said. The	rule	are ruling
(VT: sing)	are singing in the trees, they are also glad about the oasis amongst	sing	are singing
(VT: sits)	is sitting in the rocking chair and (VT: tells) is telling fairy ta	sits	is sitting
(VT: talk)	talking badly about other people at every opportunity they get. For	talk	are talking
(VT: tells)	is telling fairy tales. In summer, while people are working togethe	tells	is telling
(VT: tolerate)	are tolerating a great deal, because they (VT: depend) are depend	tolerate	are tolerating
(VT: touch)	are touching us emotionally. But in fact, they are always only a wor	touch	are touching
(VT: travel)	travelling all around the world. They are ready to enjoy everything	travel	are travelling
(VT: uses)	is using to win her interest are not convincing at all - to find unsp	uses	is using
(VT: usually get worse)	are getting worse: the law of the street is a hard one b	usually get worse	are getting
(VT: we talk)	we're talking about all the crucially important things that (VT: h	talk	are talking
(VT: wear)	am wearing latest fashion or not nor whether I do in my spare time wh	wear	am wearing
(VT: you feel strong)	you're feeling strong and sure of yourself and (VT: you ca	feel	are feeling

A 1.7 Aspectual Verb Classes

states		activities		accomplishments		achievements	
astonish	love	**attend**	roll	**attend** (a class)	obliterate	arrive	**hear**
be	**look**	continue	rotate	appoint so	paint sth	awaken	improve
believe	mean	**cook**	rumble	box	perform sth	**be** born	**kill**
belong	need	cry	**run**	buy	place sth	become	**know**
concern	own	dance	scan	bring (about)	**play** (a game)	begin	land
desire	perceive	drink	scrutinize	build sth	put	break	leave
dislike	possess	**drive**	search	cause sth VP	recover (from an illness)	catch	lose
dismay	prove	eat	seek	change (the story)	read (a book)	cease	melt
doubt	rule	gaze upon	sing	cover	rent sth	cool (down)	notice
dominate	**regret**	focus on	sit (in/on)	**cook** sth	**ride** (10 km)	cross (the border)	reach (the summit)
enjoy	**see**	follow (with the eyes)	smile	deliver sth	**run** (5 km)	darken	**see**
exist	seem		smoke	destroy		depart	start
feel	show	**go**(attend)	stay	draw sth	**run** (away)	detect	**taste**
hate	suggest	housekeep	study	**drive** (to X)	**swim** (5 km)	die	**think** of
have	smell	hum	**swim**	fly (to X)	see Carmen	discover	touch
hear	taste	keep	talk	**get** exhausted	set sth	drop	**turn** off
imply	think that	**listen** (to)	**tell** (stories)	**get** ready	shape up	end	**turn** into
involve	**understand**	**look** (for)	**think** (about)	give	take (out)	explode	spot sth
know	want	observe	vibrate	**go** to (Paris)	**tell** (a story)	fall (out)	realise
like	worry	panic	**walk**	**go** (out)	**turn** sth into sth	feel	recognise
		pay (attention)	**watch**	grow up	uncover	find	remember
		play	work	hide	**write** sth	finish	resume *V*ing
		push sth	**write** (in/on)	**kill**	**walk** (to X)	forget	**see**
		pull sth		knit sth	**watch** sth	freeze	sink
		ride (on)		make sth VP		**get** married	start *V*ing
				marry		happen	stop *V*ing
				move			**understand**
							warm (up)
							win

Aspectual verb classes after Brinton, Dowty and Collins (Brinton 1988: 241-243; Brinton 2000: 144-147; Collins 2002: 94; Dowty 1979: 66-71) – verbs in bold belong to several aspectual classes

A2 German Summary

Kapitel 1. Thema und Forschungsfrage

Die Forschung zu Aspekt hat eine lange Tradition und bietet zahlreiche theoretische Studien über die Differenzierung zwischen Aspekt als Formkategorie und als sprachliche Universalie (vgl. Andersson 1989: 29). Der Fachbegriff Aspekt – eine Lehnübersetzung aus dem Russischen *Vid* (*вид* – Ansicht, Blickrichtung) – wird häufig als Oberbegriff für die Diskussion über alle Komponenten der Aspektualität (wie z.B. morphologisch, syntaktisch, lexikalisch usw.) verwendet; dabei bildet die Kategorie des verbalen Aspekts den Kern der Diskussion. Die Kategorie des verbalen Aspekts (am Beispiel der *Progressive Form* und des *Perfect*) ist im Englischen besonders stark ausgeprägt und gehört zusammen mit ihren verschiedenen temporalen Ausprägungen (z.B. *Present Progressive* oder *Present Perfect*) zu den zentralen, aber auch zu den am schwierigsten zu erlernenden grammatikalischen Kategorien für Nicht-Muttersprachler des Englischen (vgl. Swan und Smith 1987).

Experimentelle Untersuchungen zum Gebrauch der englischen Tempora und Aspektformen zeigen, dass Lerner* des Englischen als Zweit- und/oder Fremdsprache häufig Schwierigkeiten mit dem Erlernen und Gebrauch der englischen Zeitformen aufweisen – unabhängig von der Muttersprache und dem Sprachkompetenzniveau (vgl. Shirai 2009; Salaberry 2002; Noyau 2002; Bardovi-Harlig 2000; 1999; 1994; Schumann 1987; Zydatiß 1977 usw.). Zu ähnlichen Erkenntnissen kommen auch die wenigen Studien, die als empirische Basis Sprachkorpora verwenden und den Gebrauch der englischen Zeitformen durch Lerner des Englischen erstmalig computergestützt untersuchen (vgl. Davydova 2011; Eriksson 2008, Hahn 2007, Lenko-Szymanska 2007, Schlüter 2002; Axelsson und Hahn 2001; Granger 1999, Virtanen 1997).

Obwohl kürzlich erste korpusbasierte Beschreibungen zum Gebrauch der englischen Zeitformen entstanden sind, hatte keine der bisherigen Studien den systematischen Gebrauch von Aspektformen im Englischen als Fremdsprache im Fokus und hat dabei fortgeschrittene Lerner des Englischen mit unterschiedlichen Muttersprachen verglichen. Die vorliegende Forschungsarbeit zielt darauf ab, diese Forschungslücke zu schließen. Im Mittelpunkt steht die Frage, ob und inwieweit fortgeschrittene Lerner des Englischen als Fremdsprache mit den Herkunftssprachen Deutsch und Bulgarisch die englischen Aspektformen korrekt verwenden,

und inwieweit Abweichungen von der muttersprachlichen korpusbasierten Norm festzustellen sind. Von besonderem Interesse aus linguistischer Sicht ist dabei die Frage, ob sich die Unterschiede zwischen dem englischen Aspektsystem und den Aspektsystemen der Herkunftssprachen Deutsch und Bulgarisch auf das zielsprachliche Verhalten der beiden Lernergruppen auswirken, und ob weitere Faktoren und Einflüsse eine Rolle spielen. Ferner sollen einheitliche Fehler und Probleme beim Erlernen und Gebrauch des englischen Verbsystems identifiziert werden, um in der fremdsprachlichen Vermittlung von englischen Tempus- und Aspektformen stärker adressiert werden zu können. Untersuchungsgegenstand sind Daten aus dem Lernerkorpus *ICLE* (*International Corpus of Learner English*, vgl. Granger 2009) und den muttersprachlichen Referenzkorpora *LOCNESS* (*Louvain Corpus of Native English Essays*, vgl. Granger 2002; 2009), *FLOB* und *FROWN* (vgl. Hundt et al 1998; 1999).

Die Arbeit gliedert sich in zehn Kapitel. Kapitel 2, 3 und 4 bilden die theoretischen Grundlagen mit Fokus auf Aspekt im kontrastiven Vergleich (Kapitel 2), Hypothesen zum Zweitspracherwerb und Entwicklung von Aspektualität als sprachliche Universalie (Kapitel 3) sowie den bisherigen korpusbasierten Methoden zur Untersuchung von Aspekt im Englischen als Zweit- und Fremdsprache (Kapitel 4). Im weiteren Verlauf der Arbeit werden die Korpora und die Methoden (Kapitel 5) erläutert. In der anschließenden Analyse wird zunächst eine quantitative Auswertung vorgenommen (Kapitel 6 und 7), gefolgt von einer qualitativen Fehleranalyse (Kapitel 8). Abschließend werden die Ergebnisse evaluiert und als Grundlage für integrative Modellansätze verwendet (Kapitel 9). Kapitel 10 fasst die Arbeit zusammen und bietet einen Ausblick auf zukünftige Forschung.

Kapitel 2. Theoretischer Rahmen: Aspekt im Englischen

Die Sprachen dieser Welt verfügen über diverse Mittel, um temporale Beziehungen und die Relationen zwischen ihnen auszudrücken – die Kategorie Aspekt gehört dazu. Die Forschung zu Aspekt ist Teil der Forschung über den sprachlichen Ausdruck der Temporalität und wird traditionell in zwei Hauptbereiche unterteilt: 1) die Bedeutungsbestimmung von Aspekt als Oberbegriff für die Art und Weise, wie eine Situation, ein Ereignis oder ein Zustand betrachtet wird – genannt *viewpoint aspect* oder *grammatical aspect*, und 2) die Bedeutungsbestimmung von den verschiedenen Situationstypen anhand ihrer inhärenten temporalen Eigenschaften – bezeichnet als *situation aspect*, *lexical aspect* (vgl. Smith 1983) oder auch Aktionsart (vgl. Binnick 1991: 144; Declerck 2006: 49). Die geläufigste Definition

von Aspekt im perspektivischen Sinne (*viewpoint aspect*) ist die von Comrie (1976: 3), die besagt:- „aspects are the different ways of viewing the internal temporal constituency of a situation", d.h. Aspekt schildert lediglich die Sichtweise über ein Ereignis, ohne dieses Ereignis in einer bestimmten Relation mit der Sprechzeit zu setzen. Die Opposition Perfektiv vs. Imperfektiv ist das wohl etablierteste Beispiel für *viewpoint aspect* – der perfektive Aspekt drückt abgeschlossene Ereignisse und der imperfektive Aspekt nicht-abgeschlossene Ereignisse aus. Dabei ist die Abgeschlossenheit bzw. Nicht-Abgeschlossenheit eine Eigenschaft der Zeit des Betrachtens, von der aus das Ereignis dargestellt wird. Unter Situationsaspekt oder Aktionsart hingegen versteht man die Einteilung jeder Verbalphrase nach ihren inhärenten zeitlichen Eigenschaften wie z.B. dynamisch, statisch oder zielgerichtet, und zwar unabhängig von der jeweiligen grammatikalischen Aspektmarkierung derselben Verbalphrase:

> *Aktionsart*, in its narrower sense, relating only to the lexical verb in question, provides lexical information. The arguments and adjuncts of the verb may provide further information [...] on the context of the situation, or more generally pragmatics. (vgl. Comrie 2001: 43)

Obwohl sich die Sprachwissenschaft über diese viel diskutierte Unterscheidung uneinig ist, wird Aktionsart traditionell als der lexikalische Ausdruck der semantischen Kategorie Aspektualität verstanden und Aspekt als der grammatikalische, morphologisch markierte Ausdruck verwendet (vgl. Schüller 2005). Die gängigste Kategorisierung von Aktionsarten ist die von Vendler (1957), der zwischen *states* (statisch, andauernd und atelisch), *activities* (dynamisch, andauernd und atelisch), *accomplishments* (dynamisch, andauernd, telisch) und *achievements* (dynamisch und punktuell) unterscheidet (vgl. auch Andersen und Shirai 1996). Diese Taxonomie basiert auf der sogenannten Telizität oder der Eigenschaft einer Situation „ein bestimmtes Ziel oder Grenzwert zu erreichen, bei dem die Handlung aufhört oder in eine andere Handlung übergeht" (vgl. Andersen 1972 in Abraham 1989: 9). Der Zusammenhang und die Interaktion zwischen Aspekt und Aktionsart ist umstritten und dient als Grundlage für die Aspekttheorie (siehe Kapitel 3).

Ähnlich umstritten ist die Abgrenzung zwischen Aspekt und Tempus – als Hauptunterscheidungsmerkmal gilt dabei die Eigenschaft, Situationen oder Ereignisse in Relation zu der Sprechzeit zeitlich zu lokalisieren. Traditionell bezieht sich Tempus auf den Sprechzeitpunkt und ist daher eine deiktische Kategorie, während Aspekt nicht-deiktisch ist, da er eine Sichtweise über das Ereignis ausdrückt (z.B. abgeschlossen oder nicht-abgeschlossen), ohne dieses Ereignis in Verbindung mit der tatsächlichen Sprechzeit zu

setzen. Die klassische Tempustheorie von Hans Reichenbach (1952) definiert drei Zeitpunkte – die Sprechzeit **S**, die Ereigniszeit **E** und die Referenzzeit **R**, von der aus das Ereignis betrachtet wird. Klein (1994) hingegen bietet ein überarbeitetes Modell von Reichenbachs Tempustheorie, in dem er die Parameter *Time of Utterance* (**TU**, parallel zur deiktischen Kategorie **S**), *Time of Situation* (**TSit**, parallel zu **E**) und *Time of Assertion* (**T-ASS**) als eine Präzisierung von Reichenbachs Referenzzeit **R** verwendet. **T-ASS** bezeichnet eine Zeitspanne, für die die Äußerung des Sprechers gilt. Demnach ist Tempus eine temporale Beziehung zwischen **TU** und **T-ASS**, während Aspekt die temporale Beziehung zwischen **T-Sit** und **T-ASS** ausdrückt (vgl. Klein 1995: 143). Kleins Theorie zufolge gibt es die folgenden Tempora und Aspekte im Englischen:

TENSE		ASPECT	
FUTURETU before	T-ASS	IMPERFECTIVE	T-ASS in T-SIT
PRESENT	TU INCL T-ASS	PERFECTIVE T-ASS	OVL T-SIT and TIME after T-SIT
PAST	TU AFTER T-ASS	PERFECT	T-ASS AFTER T-SIT
		PROSPECTIVE	T-SIT AFTER T-ASS

Tempora und Aspekte im Englischen (vgl. Klein 1995: 144)

Andere Theorien wie z.B. die von Comrie (vgl. Comrie 1976:3; 34) nehmen an, dass die Opposition *progressive – non-progressive* im Englischen die geläufigste aspektuelle Unterscheidung darstellt, wobei diese Unterscheidung als ein Sonderfall gilt und der Imperfektiv-Kategorie zuzuordnen ist. Von besonderem linguistischem Interesse ist zudem der Status des englischen *Present Perfect*, das abwechselnd als Aspekt, relatives Tempus (vgl. Binnick 1991) oder eine dritte Kategorie (vgl. Bybee 1994; Kortmann 1995) definiert wird. Entscheidend für die aspektuelle Bedeutung des *Present Perfects* ist die „Verknüpfung einer in der Vergangenheit angelegten Situation mit der Gegenwart oder mit einer Zeit, die zumindest nach der betrachteten Situation liegt […] was als aspektuelle Komponente betrachtet wird" (vgl. Schüller 2005: 34; Kortmann 1995: 185). Die vorliegende Arbeit richtet sich nach jenen Theorien und Referenzwerken, die das *Present Perfect* als einen aspektuellen Sonderfall behandeln (vgl. Comrie 1976; Quirk et al. 1985; Biber et al. 1999)

Die Beschreibung des Aspekts in wissenschaftlichen Referenzgrammatiken wie Quirk et al. (1985) oder Biber et al. (1999) unterscheidet sich wesentlich „von der Thematisierung desselben Gegenstands in theoretischen Abhandlungen" (vgl. Schüller 2005: 29). Aspekt als formale Kategorie ist "the grammatical category which reflects the way in which the verb action is regarded or experienced with respect to time" (Quirk et al. 1985: 188) und wird im Englischen mit den Oppositionen 1) *Progressive Aspect* (*He was reading*) vs. *Simple Aspect*

(*He read*) und 2) *Perfect Aspect* (*He has read*) ausgedrückt, wobei der *Simple Aspect* die überwiegende Variante darstellt (vgl. Biber 2006: 63). Die wichtigste Funktion des *Progressive Aspect* ist die des Verlaufs bzw. die Bezeichnung von temporären Situationen, die zeitlich eine begrenzte Gültigkeit haben und nicht unbedingt als abgeschlossen gelten (vgl. Quirk et al. 1985: 198). Weitere Funktionen sind die der begrenzten Dauer, die Habitualität (oft in Zusammenhang mit Adverbien wie *usually, always* usw.), der Ausdruck von Emotionen oder Höflichkeitsformen usw. (vgl. Mindt 2000: 256 – 261). Der *Progressive Aspect* wird meistens mit Tätigkeitsverben (*activities*) oder Vorgangsverben (*accomplishments*) kombiniert und selten mit statischen Verben – insbesondere mit Verben der passiven Wahrnehmung und Kognition (z.B. *smell, understand* usw.) oder Habens und Seins (*be, belong, own* usw.), die als unzulässig gelten (vgl. Quirk et al. 1985: 200-201). Das *Present Perfect* hingegen setzt einen Zustand in Beziehung zu einem vorhergehenden Ereignis (vgl. Comrie 1976: 52) und hat vier Hauptbedeutungskomponenten: 1) *indefinite past* (*resultative* und *non-resultative*), 2) *continuative past*, 3) *recent past* und 4) *completion* (vgl. Mindt 2000: 224), wobei manche Bedeutungskomponenten wie z.B. die Funktion des *Present Perfects* als unbestimmte Vergangenheit (resultativ und nicht-resultativ) viel häufiger auftreten.

Darüber hinaus sind die Angaben zur Auftretenshäufigkeit des *Progressive Aspect* und des *Perfect Aspect* als Formkategorien von besonderer Bedeutung für die vorliegende Arbeit, da die Arbeit Lernersprache korpusbasiert untersucht und die Häufigkeitswerte der Lerner mit den von Referenzgrammatiken festgestellten Häufigkeitswerten vergleicht, um mögliche Abweichungen von der korpusbasierten muttersprachlichen Norm untersuchen zu können. Biber et al. (1999: 461) stellen fest, dass der *Simple Aspect* die überwiegende Variante in allen Sprachregistern des heutigen Englisch (z.B. Sachtexte, gesprochene Sprache, Belletristik usw.) darstellt, gefolgt von dem *Perfect Aspect* und *Progressive Aspect* (siehe Grafik). Dabei ist die *Progressive* Form häufiger im amerikanischen Englisch, während die *Perfect* Form häufiger im britischen Englisch vorkommt (vgl. Biber et al. 1999: 461 – 462).
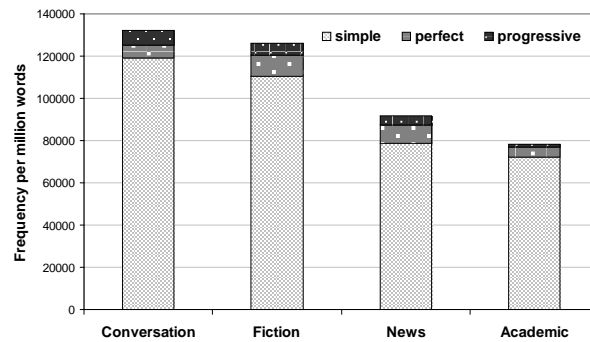
Relative Häufigkeiten der Progressive, Perfect und Simple Form (vgl. Biber et al. 1999: 461)

Das wesentliche Unterscheidungsmerkmal zwischen der Zielsprache Englisch und den Herkunftssprachen der beiden Lernergruppen Deutsch und Bulgarisch ist die Ausprägung der Aspektualität als semantische Kategorie bzw. die Frage, welche grammatikalischen und/oder lexikalischen Mittel dafür verwendet werden, um die Bedeutung des Englischen *Progressive* und *Perfect Aspect* im Deutschen oder Bulgarischen entsprechend auszudrücken. Diese Unterschiede wirken sich möglicherweise auf den Zweitspracherwerb bzw. den Gebrauch der englischen Aspektformen der beiden Lernergruppen in der vorliegenden Arbeit aus.

Eine grammatikalisierte Aspektopposition im klassischen Sinne zwischen *Progressive* und *Non-Progressive* Verbformen gibt es im Deutschen nicht – Progressivität bzw. Imperfektivität wird daher mit anderen sprachlichen Mitteln ausgedrückt (vgl. Filip 1989; Ebert 2000; Königs 1995; König und Gast 2009 usw.). Traditionell wird das englische Progressiv mit unterschiedlichen Adverbien im Deutschen übersetzt bzw. wiedergegeben (z.B. *gerade, nun, jetzt, zurzeit, momentan,* usw.); jedoch gewinnen Konstruktionen wie *am+Vinf* (z.B. *am arbeiten*) zunehmend an Bedeutung und deuten eine allmähliche Grammatikalisierung bzw. Progressiv-Markierung außerhalb des Rhein-Ruhr Dialekts an. König and Gast (2009: 93) verwenden die folgende Skala, um die Häufigkeit dieser Konstruktionen aufzuzeigen:

$$am + V_{\text{nominalisiert}} > dabei + \text{infinitiv} > beim + V_{\text{nominalisiert}} > im + V_{\text{nominalisiert}}$$

Im Gegensatz zur fehlenden Progressiv-Kategorie im Deutschen ist die Kategorie des Perfekts vollständig grammatikalisiert (vgl. Löbner 2002: 373). Das Perfekt im Deutschen hat eine hohe formale Ähnlichkeit zu dem englischen Perfekt (*haben* + Partizip II), fungiert aber im Gegensatz zum englischen *Present Perfect* zunehmend als analytische Tempusform der Vergangenheit mit deiktischer Interpretation. Somit hat das Perfekt in vielen regionalen

Varietäten des Deutschen das Präteritum fast vollständig ersetzt (Oberdeutscher Präteritumschwund), da die beiden Formen in ihrer temporellen Bedeutung als semantisch (aber nicht stilistisch) austauschbar gelten (vgl. Löbner 2002; Klein 2000; Klein und Vater 1998). Dazu kann das Perfekt im Deutschen mit vergangenheitsbezogenen Adverbien (z.B. *gestern, letztes Jahr* usw.) beliebig kombiniert werden. Eine weitere Besonderheit des deutschen Perfekts ist seine Mehrdeutigkeit – zusätzlich zu seiner temporellen Bedeutung hat das deutsche Perfekt auch eine aspektuelle Funktion der Abgeschlossenheit (ähnlich wie das englische resultative *Perfect*), die z.B. nicht durch das Präteritum ausgedrückt werden kann (vgl. König and Gast 2009: 86): „Wir brauchen Hilfe – unser Hund ist weggelaufen! vs. *Wir brauchen Hilfe – Unser Hund lief weg!"

Die formale Ähnlichkeit des deutschen Perfekts zum englischen Perfect ist insofern problematisch, dass deutsche Lerner des Englischen als Fremdsprache möglicherweise die Funktionen des englischen *Present Perfect* mit denen des deutschen Perfekts verwechseln und es dann fälschlicherweise als Tempus der Vergangenheit im Englischen verwenden wollen.

Demgegenüber unterscheidet sich das bulgarische Aspektsystem sowie die bulgarische Sprache als eine südslawische Sprache wesentlich vom englischen Aspektsystem. Dabei verfügt das Bulgarische über ein sehr differenziertes Aspektsystem nach dem slawischen Muster (vgl. Aronson 1984; Dahl 2000; Lindstedt 1985; Scatton 1984): es gibt die klassische systematische Aspektopposition zwischen dem perfektiven (vollendeten) und imperfektiven (unvollendeten) Aspekt sowie eine Unterscheidung zwischen Imperfekt und Aorist (nicht-abgeschlossene vs. abgeschlossene Vergangenheit). Die imperfektiven-perfektiven Opposition im Bulgarischen wird durch Präfigierung, Suffigierung oder Suppletivismus ausgedrückt und wird daher von manchen Forschern als morphologisch-lexikalisch empfunden und der Kategorie Aktionsart zugeordnet (vgl. Bertinetto und Delfitto 2000: 190). Imperfektive und perfektive Verben bilden dabei ein Paar und werden meistens voneinander abgeleitet (der perfektive Wortstamm bildet dabei die markierte Form). Die zweite Opposition zwischen den synthetischen Formen Aorist vs. Imperfekt wird auch als aspektuell empfunden, da der Aorist dazu dient, abgeschlossene Ereignisse aus der Vergangenheit einfach zu konstatieren (auch Erzählform genannt), während das Imperfekt Handlungen oder Ereignisse im Verlauf zeitlich lokalisiert, wobei der Beginn oder das Ende der Handlung außerhalb des Blickfelds liegen (vgl. Radeva 2003). Die größte funktionale Ähnlichkeit zum Englischen *Progressive Aspect* stellt daher das Imperfekt im Bulgarischen: so kann das englische *Past*
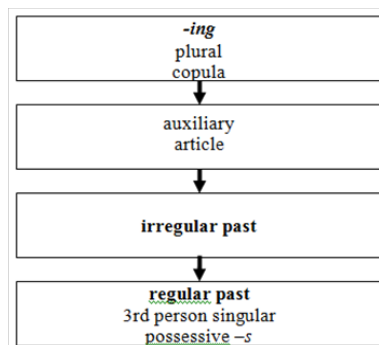
Progressive am besten mit dem bulgarischen Imperfekt übersetzt bzw. wiedergegeben werden, während das *Simple Past* eher mit dem Aorist wiedergegeben werden kann (vgl. Scatton 1984: 322 – 323). Das bulgarische Perfekt ist hingegen eine analytische Form, die dazu verwendet wird, Handlungen auszudrücken „deren Stattfinden in der Vergangenheit liegt, die Tatsache ihres Stattfindens jedoch für die Gegenwart von Bedeutung ist" (vgl. Radeva 2003: 120). Demzufolge hat das bulgarische Perfekt, ähnlich wie das englische *Present Perfect* in seiner indefiniten, nicht-resultativen Funktion (*Experiential Perfect*) die Funktion des Erfahrungsperfekts behalten. Darüber hinaus hat das bulgarische Perfekt eine weitere nicht-aspektuelle Sonderfunktion: die der Evidentialität oder des Ausdrucks von Wissen „zweiter Hand". So fungiert das bulgarische Perfekt als indirekte Erzählform (Renarrativ, nacherzählte Formen) von Ereignissen, die man nicht persönlich erlebt hat, oder von deren Informationsquelle man sich distanzieren möchte (vgl. Bybee 1985; Lindstedt 2000; Radeva 2003). Somit wird zusätzlich epistemische Modalität durch das bulgarische Perfekt zum Ausdruck gebracht.

Der kontrastive Vergleich zwischen Englisch als Zielsprache und Deutsch und Bulgarisch als Muttersprachen dient dazu, den Gebrauch der englischen Aspektformen durch fortgeschrittene Lerner des Englischen mit den beiden Muttersprachen besser erklären zu können, insbesondere in Bezug auf bestimmte Charakteristika der Lernergruppen oder mögliche Abweichungen von der korpusbasierten muttersprachlichen Norm.

Kapitel 3. Aspekt im Zweitspracherwerb

Das vorliegende Kapitel widmet sich den Zweitspracherwerbstheorien, die sich mit dem Erwerb von Temporalität und insbesondere von Aspekt als Ausdruck der Temporalität beschäftigen. Im speziellen werden hier die zwei Hauptstränge in der Zweitspracherwerbsforschung von Temporalität besprochen – der formorientierte Forschungsansatz (*form-oriented approach*) und der konzeptorientierte Forschungsansatz (*concept-oriented approach*). Der formorientierte Forschungsansatz fokussiert auf das Vorkommen von bestimmten Formen in der Lernersprache (wie z.B. Tempus- oder Aspektformen), während der konzeptorientierte Forschungsansatz alle möglichen Formen des Ausdrucks der temporalen Lokalisierung untersucht (vgl. Bardovi-Harlig 2007; 1999; Dietrich, Klein and Noyau 1995; Meisel 1987; von Stutterheim and Klein 1987 usw.). Letzteres steht nicht im Fokus der vorliegenden Studie, da sich diese nach dem Vorkommen bestimmter Formen in der Lernersprache richtet.

Die frühe formorientierte Zweitspracherwerbsforschung – die sogenannten *Morpheme Order Studies* – widmet sich unter Anderem der Reihenfolge im Zweitspracherwerb von Verbflexionen (vgl. Dulay and Burt 1974; Krashen 1978; Gass 1994; Dulay 1974 usw.). Ihre wichtigsten Ergebnisse sind, dass Lerner des Englischen mit unterschiedlichen Muttersprachen meistens die gleiche Erwerbsabfolge von bestimmten Morphemen aufweisen (z.B. das *ing*-Morphem wird als erstes erworben) – die sogenannte *Natural Order* Erwerbsabfolge (vgl. Krashen 1978: 190 – siehe Grafik):

```
         -ing
        plural
        copula
           |
           v
       auxiliary
        article
           |
           v
      irregular past
           |
           v
       regular past
    3rd person singular
       possessive -s
```

Krashens Natural Order (vgl. Krashen 1978: 190)

Spätere formorientierte Studien bestätigen die Ergebnisse der früheren *Morpheme Order Studies*, stellen jedoch fest, dass es auch individuelle Unterschiede, insbesondere in Bezug auf den gleichzeitigen Erwerb der dazugehörigen Funktionen gibt (*Form-Function Mapping*) (vgl. Bardovi-Harlig 1997; 2000; Dietrich, Klein und Noyau 1995; Housen 2002a; 2002b usw.). Demnach wird das *Bare Progressive* als erstes erworben, und die vollständige *Perfect* Form erst in der vierten Phase des Zweitspracherwerbs nach Housen 2002a; 2002b (siehe Grafik).

Stage	Category	Comment	Example
0	Invariant verb form	V_0	*see, play*
1	Present participle **Ving** Irregular past of be	Initially without aux. *be*	*seeing, playing* *was*
2	Irregular past of other verbs		*had, got*
3	Regular past *Ved* Future be-going-to + *Vinf*	Allomorphs: without aux. *be, to, -ing, gonna*	*played, worked* *is going married*
4	**Perfect aux + *V***	Allomorphs: aux. *be* and *have*; initially $V = V_0$	*have see* *is fall* *is fallen* *has fall* *have fallen*
	Present *Vs* Future will + *V*		*goes, comes,* *will make, will see*

Erwerbsabfolge von Tempus- und Aspektformen im Englischen (vgl. Housen 2002b: 158)

Zusätzlich zu den formorientierten Studien zum allgemeinen Tempus- und Aspekterwerb und Gebrauch wurden auch zwei wichtige Hypothesen aufgestellt, die die Erwerbsabfolge von Verbflexionen im Zweitspracherwerb als universales Lernerverhalten,

unabhängig von der Muttersprache darstellen: die Aspekthypothese (*Aspect Hypothesis*, AH) und die Diskurshypothese (*Discourse Hypothesis*, DH). Die Aspekthypothese basiert auf der frühen Annahme, dass Aspekt im Erst- sowie Zweitspracherwerb vor Tempus erworben wird (*Aspect before Tense*, auch *Defective Tense Hypothesis*) und dass Fremdsprachenlerner jeden Alters und Kinder beim Erstspracherwerb gleichermaßen Verbflexionen dazu nutzen, um die inhärenten Aktionsarten nach Vendler (1957) fälschlicherweise zu markieren (vgl. Vendler 1957; Weist, Wysocka et al. 1984; Robison 1995; Andersen und Shirai 1996; Bardovi-Harlig 1994; 1999; 2000; Rohde 2002; Shirai 2007 usw.). Somit werden:

1) alle telischen Verben (*accomplishments* und *achievements*) zuerst mit perfektiven Verbflexionen markiert; später bekommen auch statische Verben und Tätigkeitsverben (*activities*) ebenfalls perfektive Markierungen
2) die imperfektive Vergangenheit wird nach der perfektiven Vergangenheit erworben
3) in Sprachen, die über die Kategorie *Progressive* verfügen, werden zuerst Tätigkeitsverben (*Activities*) für das Progressiv markiert, gefolgt von *Accomplishments* und *Achievements*
4) Progressiv-Markierungen werden nicht auf statische Verben falsch übertragen (vgl. Andersen and Shirai 1996: 533)

Tatsächlich wurden auch ähnliche Tendenzen in dem muttersprachlichen Input festgestellt – erwachsene Muttersprachler weisen ebenfalls eine ungleiche Verteilung der Verbflexionen in ihrer Sprache auf, so dass z.B. telische Verben häufiger perfektive Markierungen bekommen. Dieses Sprachverhalten – bekannt auch als *Distributional Bias* (vgl. Andersen und Shirai 1995; 1996) spiegelt sich in der Lernersprache wieder; dabei ist die Ausprägung bei Lernern und Muttersprachlern unterschiedlich stark: „the exact pattern [of distribution of verbal morphology] will vary depending on L1, L2 [and] […] verbal morphology correlates with lexical aspect at least during some stage in the development of an interlanguage" (vgl. Robison 1990: 330).

Die zweite Hypothese – die Diskurshypothese – geht von einem Zusammenhang zwischen Aspektmarkierungen und Erzählvorder- und Hintergrund aus (z.B. Godfrey 1980, Givon 1987; 1989; Hopper 1989; Noyau 2002 usw.). Demzufolge wird der Diskursvordergrund durch sprachliche Mittel deutlich von dem Diskurshintergrund getrennt, so dass im Vordergrund hauptsächlich Verben mit perfektivischer Markierung erscheinen, während die Verben mit imperfektivischer Markierung im Hintergrund bleiben. Wenn in dem Diskursvordergrund hauptsächlich telische Verben mit perfektivischer Markierung und in dem Hintergrund atelische Verben mit imperfektivischer Markierung vorkommen, überlappen die beiden Hypothesen.

Neben dem von den beiden Hypothesen aufgestellten universalen Lernerverhalten spielt der Einfluss der Muttersprache ebenfalls eine wichtige, wenngleich umstrittene Rolle. Collins (2002; 2004), Rohde (2002) und Odlin (2008) plädieren für einen gemäßigten Einfluss der Muttersprache auf den Erwerb und Gebrauch von Tempus und Aspekt, der die Aspekthypothese gegebenenfalls verstärken oder abschwächen kann (vgl. Rohde 2002: 211) – der sogenannte "developmentally constrained L1 influence" (Collins 2004: 254). Andere Studien wie Slabakova (2000) und Shirai und Nishi (2003) stellen dagegen fest, dass sich Unterschiede zwischen den Muttersprachen der Lerner und der Zielsprache wie z.B. Telizität auf das Lernerverhalten negativ auswirken, insbesondere wenn die Lerner noch Anfänger sind. Im nächsten Schritt wird nun der Frage nachgegangen, inwieweit der Fortschritt im Zweitspracherwerb von Temporalität auch korpusbasiert und kontrastiv untersucht werden kann.

Kapitel 4. Lernerkorpora in der Zweitspracherwerbsforschung zu Aspekt

Die vorliegende Dissertation basiert auf einer computergestützten korpuslinguistischen Untersuchung von authentischer Lernersprache in Schriftform und gliedert sich somit in die Lernerkorpusforschung an der Schnittstelle zur Fremdsprachen- und Grammatikforschung ein. Die korpusbasierte Grammatikforschung zählt zu den neuesten Entwicklungen der Korpuslinguistik und zielt darauf ab, auf der Grundlage von Sprachkorpora (digitalisierte Sammlungen authentischer Schriftsprache oder gesprochener Sprache) bestimmte Hypothesen über den tatsächlichen Grammatikgebrauch zu überprüfen, zu bestätigen oder zu widerlegen. Der Nutzen von Korpora in der Grammatikforschung liegt darin, dass die in den Korpora enthaltenen Daten ihren Ursprung in authentischen schriftlichen oder mündlichen Äußerungen haben, die mit einem bestimmten Forschungsziel unter Berücksichtigung bestimmter Kriterien gesammelt wurden, maschinenlesbar und quantifizierbar sind. Die Hauptvorteile der korpusbasierten Grammatikforschung sind die Feststellung von Häufigkeiten und die automatisierte Erkennung von grammatikalischen Strukturen wie z.B. Häufigkeiten von Morphemen und Wortarten und ihre Verteilung in verschiedenen Sprachvarietäten und Gattungen, Häufigkeiten von syntaktischen Konstruktionen usw. (vgl. Meunier 2007: 25; Biber, Conrad und Reppen 1998: 57-58). Die Ergebnisse der ersten umfangreichen Bestrebungen zur korpusbasierten Beschreibung der Grammatik des heutigen Englisch liegen in Referenzgrammatiken wie die *Longman Grammar of Spoken and Written English* (Biber et al. 1999) oder *An Empirical Grammar of the English Verb System (vgl.* Mindt 2000) vor.

Ebenfalls zu den jüngsten Entwicklungen der Korpuslinguistik zählt die Lernerkorpusforschung, basierend insbesondere auf fortgeschrittenen Lernervarietäten. Korpuslinguistische Zugänge zur Lernersprache ermöglichen eine detaillierte Analyse und Beschreibung der Lernersprache als Interimssprache (vgl. Selinker 1972) und helfen dabei, typische Muster einer bestimmten L2-Erwerbsphase oder einer bestimmten Lernergruppe herauszufiltern bzw. das nicht-muttersprachliche Lernverhalten quantitativ und qualitativ auszuwerten (vgl. Gries 2008; Granger 1996; 1998; Granger et al. 2002; 2004a; 2004b usw.). Als Grundlage der Analyse dienen Lernerkorpora oder Sprachsammlungen authentischer L2 Sprachdaten, die Granger folgendermaßen definiert:

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance. (vgl. Granger 2002: 7)

Die existierenden Lernerkorpora haben sich in dem vergangenen Jahrzehnt als wertvolle Ressource für eine Reihe von empirischen Untersuchungen etabliert und dienen gleichermaßen der Sprachwissenschaft und des Fremdsprachenunterrichts (Granger et al. 2002; 2004a; 2004b; Mukherjee 2006b; 2006c; Pravec 2002 usw.). Besonders aufschlussreich sind Lernerkorpora, die Sprachdaten unterschiedlicher Lernervarietäten enthalten (Multi-L1 Lernerkorpora), da sie breitaufgestellte Analysen des allgemeinen Lernerverhaltens vornehmen und auf die gemeinsamen Fehler z.B. im lexikalischen Bereich fokussieren können (vgl. Granger et al. 2002; 2004). Darauf basierend können Unterrichtswerke wie Wörterbücher (z.B. das *Macmillan English Dictionary*, vgl. De Cock et al. 2007) überarbeitet werden (*delayed pedagogic purposes*), um die problematischen Bereiche in der Fremdsprache stärker adressieren zu können. Im ersten Schritt wird den folgenden Fragen nachgegangen (vgl. Leech 1998: xiv):

1) Welche linguistischen Elemente (z.B. Wörter, Wortarten, syntaktische Konstruktionen usw.) werden von Lernern im Vergleich zu den muttersprachlichen Referenzkorpora zu häufig (*overuse*, Übergebrauch) oder zu selten (*underuse*, Mindergebrauch) verwendet?
2) Welche Elemente werden gar nicht verwendet (*avoidance*)?
3) Welche sprachlichen Unterschiede gibt es zwischen den Lernergruppen mit unterschiedlichen Muttersprachen?
4) In welchen Bereichen erreichen Lerner ein zielsprachliches Verhalten?
5) Welche Bereiche der Zielsprache sind am schwierigsten?

Diese Fragen stellen die Grundlage des kontrastiven Forschungsansatz zur Analyse von Lernersprache (*Contrastive Interlanguage Analysis*, vgl. Granger 1996; 1998; Gilquin 2008 usw.) dar – eine der zwei Hauptmethoden in der Lernerkorpusforschung, die mit Hilfe

von statistischen Verfahren Lernertexten mit muttersprachichen Texten vergleicht, um Abweichungen von der korpusbasierten muttersprachlichen Norm festzustellen. Die zweite Methode ist die computergestützte Fehleranalyse (*Computer-Aided Error Analysis* vgl. Dagneaux et al. 1998), die Fehler annotiert, klassifiziert und quantifiziert, um die Ursprünge des nicht-zielsprachlichen Lernerverhaltens besser erklären zu können. Das übergeordnete Ziel ist es, die durch korpuslinguistische Methoden gewonnenen Erkenntnisse im Fremdsprachenunterricht anwenden zu können, um Entscheidungen über die Relevanz gewisser Unterrichtsthemen besser treffen zu können (siehe Grafik).
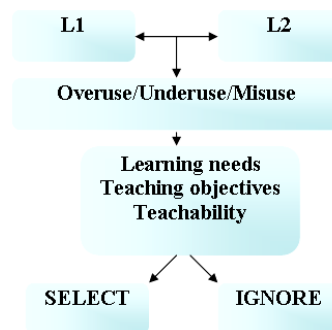
Entscheidungsschema im Fremdsprachenunterricht (vgl. Granger 2009: 23)

Obwohl die Mehrzahl der bisherigen Korpusstudien Lernersprache auf der lexikalischen Ebene untersucht, gibt es bereits einige Studien, die den Gebrauch von Tempus- und Aspektformen durch fortgeschrittene Lerner des Englischen als Zweit- und Fremdsprache im Fokus haben, jedoch nicht mit dem englischen *Progressive* und *Perfect* als Hauptuntersuchungsgegenstand (vgl. Virtanen 1997; Granger 1999; Axelsson und Hahn 2001; Housen 2002a; 2002b; Lenko-Szymanska 2007; Eriksson 2008; Wulff und Römer 2009; Hundt und Vogel 2011; Davydova 2011). All diese Forschungsarbeiten kommen zu dem Schluss, dass auch sehr fortgeschrittene Lerner des Englischen die englischen Tempus- und Aspektformen in der Schriftsprache fehlerhaft verwenden, obwohl sie zu unterschiedlichen Ergebnissen in Bezug auf den Minder- und Übergebrauch und den Einfluss der Muttersprache kommen. Beim Gebrauch des *Progressive Aspects* durch deutsche, schwedische, finnische, französischsprachige, niederländischsprachige und polnische Lerner des Englischen werden Abweichungen von der muttersprachlichen korpusbasierten Norm wie z.B. leichter Übergebrauch bei einigen dieser Lernergruppen festgestellt. Viel interessanter sind jedoch die Abweichungen von einer bestimmten muttersprachlichen Varietät wie z.B. amerikanisches Englisch oder die mehrdeutigen Fälle, die keine „richtigen" Tempus- und Aspektfehler darstellen, aber möglicherweise auf den Einfluss der Muttersprache zurückzuführen sind. Zusätzlich wird die Aspekthypothese in ihrer vierten Aussage widerlegt,

da Lerner das *Progressive Aspect* auch auf statische Verben inkorrekt übertragen (vgl. Housen 2002a; 2002b; Wulff und Römer 2009). Darüber hinaus zeigen Granger (1999) und Eriksson (2008) auf, dass Lerner häufig Schwierigkeiten mit dem richtigen Gebrauch von Tempus- und Aspektformen im Kontext haben; dazu stellen die Formen und Funktionen des *Present Perfect* besondere Probleme für die Lerner dar und werden gar gemieden (vgl. Davydova 2011). Da sich insbesondere der vollständige Erwerb der englischen Aspektformen auch in den späten Phasen des Zweitspracherwerbs als schwierig erweist (vgl. Meunier und Littre 2013), geht die vorliegende Dissertation der Frage nach, inwieweit dies für fortgeschrittene deutsche und bulgarische Lerner zutrifft und inwieweit Unterschiede zwischen den beiden Lernergruppen festgestellt werden können.

Kapitel 5. Korpora und Methoden

Im Fokus der vorliegenden Dissertation steht die kontrastive Analyse zwischen zwei fortgeschrittenen Lernervarietäten des Englischen als Fremdsprache – die Lernervarietät bulgarischer Lerner des Englischen und die Lernervarietät deutscher Lerner des Englischen. Der zugrunde liegende Forschungsansatz ist eine erweiterte Form der obenerwähnten *Contrastive Interlanguage Analysis* (vgl. Granger 1996; 1998), die darauf abzielt, die zwei Lernervarietäten mit vier unterschiedlichen Referenzvarietäten des Englischen zu vergleichen – einerseits mit britischem und amerikanischem Englisch und andererseits mit dem Schriftenglisch von Muttersprachlern, die Schreibnovizen und Schreibexperten sind (siehe Grafik).
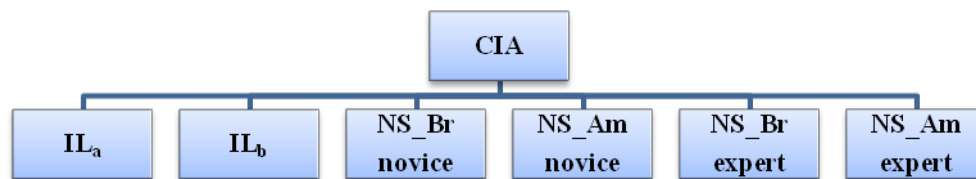
Erweiterte Form der Contrastive Interlanguage Analysis

Ziel des erweiterten Vergleichs ist es, eventuelle Gemeinsamkeiten zwischen dem Aspektgebrauch in der Lernersprache mit dem Aspektgebrauch in einer bestimmten muttersprachlichen Varietät (z.B. amerikanisches Englisch) feststellen zu können, um der Frage der expliziten und impliziten fremdsprachlichen Zielnorm für die jeweilige Lernergruppe besser nachgehen zu können. Ferner ist es von linguistischem Interesse, ob und inwieweit das grammatikalische Sprachverhalten der deutschen und bulgarischen Lerner des Englischen dem Sprachverhalten britischer und amerikanischer Schreibnovizen ähnelt bzw.

ob die Lerner- und muttersprachlichen Varietäten ein Kontinuum in Bezug auf konzeptionelle Schriftlichkeit bilden (vgl. Koch und Oesterreicher 1985).

Die Software Tools, die für den quantitativen Vergleich herangezogen wurden, sind *WordSmith* Tools Version 4 (vgl. Scott 2010) und *Wmatrix* (vgl. Rayson 2008). *WordSmith Tools* ist eine Standardkorpussoftware für die statistische Auswertung von z.B. Worthäufigkeiten, während *WMatrix* eine webbasierte Korpussoftware ist, die über zusätzliche Funktionen wie z.B. Wortartkategorisierung (*Tagging*) verfügt. Der Vorteil von Wortartkategorisierung liegt darin, dass alle Verben mit *WMatrix* mit *Part-of-Speech Tags* kategorisiert und die finiten Verben folglich extrahiert werden können.

Als Datenbasis für die Untersuchung dient das Lernerkorpus *ICLE* (*International Corpus of Learner English*) in seiner überarbeiteten Version (vgl. Granger et al. 2002; 2009). *ICLE* ist das bisher bekannteste Lernerkorpus des Englischen als Fremdsprache und beinhaltet ca. 3,7 Mio. Wörter in 16 Komponenten mit Lernertexten von Lernern mit 16 unterschiedlichen Muttersprachen. Die Texte in *ICLE* stammen von fortgeschrittenen Studenten der Anglistik (im zweiten und dritten Studienjahr) und basieren hauptsächlich auf Essays, die zu einem kontroversen Thema im Unterrichtskontext oder als Hausaufgabe geschrieben wurden (z.B. *In his novel "Animal Farm" George Orwell wrote "All men are equal but some are more equal than others". How true is this today*?). Für die vorliegende Arbeit wurden zwei Korpora von der deutschen und bulgarischen Komponente des *ICLE*– *BUCLE* und *GICLE* extrahiert. Die beiden Lernerkorpora beinhalten nur Argumentationsaufsätze geschrieben von deutschen und bulgarischen Muttersprachlern und sind folgendermaßen aufgebaut (siehe Tabelle):

BUCLE	Universität Sofia	**300** Aufsätze	**199.249** Wörter	**664,16** Aufsatz-durchschnittslänge	**99.5 %** der Gesamtkomponente
GICLE	Universitäten Augsburg, Basel, Dresden, Salzburg usw.	**429** Aufsätze	**226.503** Wörter	**527,97** Aufsatz-durchschnittslänge	**96 %** der Gesamtkomponente

Lernerkorpus Design – *BUCLE* und *GICLE*

Um Vergleichbarkeit zu gewährleisten, wurden die Komponenten nach den gleichen Vorgaben kompiliert; dennoch ist *BUCLE* etwas homogener als *GICLE*. Darüber hinaus unterscheiden sich die beiden Lernerkorpora in Bezug auf lernerbezogene bzw. außersprachliche Parameter wie z.B. den Auslandsaufenthalt bzw. den Kontakt mit der Zielsprache Englisch (über 60% der deutschen Lerner waren bereits mindestens für einen

Monat im englischsprachigen Ausland im Vergleich zu weniger als 10% der bulgarischen Lerner) oder das Kompetenzniveau (mehr *GICLE* Lerner als *BUCLE* Lerner wurden von unabhängigen Gutachern in das höchste Kompetenzniveau C2 des Gemeinsamen Europäischen Referenzrahmen eingestuft, vgl. Granger et al. 2009).

Die muttersprachlichen Referenzkorpora basieren einerseits auf dem Pendant von *ICLE* – dem *LOCNESS* Korpus (*Louvain Corpus of Native English Essays*), einem Korpus mit Schüler- und Studentenaufsätzen mit ca. 324,304 Wörtern und britischer (*LOCNESS_br*) und amerikanischer (*LOCNESS_us*) Komponente, und andererseits auf Standardreferenzkorpora des britischen und amerikanischen Englisch wie *FLOB* (*Freiburg LOB Corpus of British English*) und *FROWN* (*Freiburg BROWN Corpus of American English*, vgl. Hundt et al. 1999). Die vier muttersprachlichen Korpora beinhalten nur nicht-technische argumentative Texte – studentische Aufsätze zu kontroversen Themen aus *LOCNESS* und veröffentlichte argumentative Beiträge aus Zeitungen und Zeitschriften von *FLOB* und *FROWN* – die F Komponenten *FLOB_F* und *FROWN_F*. Die Hauptunterschiede zwischen den Referenzkorpora liegen darin, dass 1) sie zwei Varietäten des Englischen – britisches und amerikanisches Englisch abbilden, und 2) Texte von muttersprachlichen Schreibnovizen (die *LOCNESS* Teilkorpora) und muttersprachlichen Schreibexperten (die *FLOB* und *FROWN* Teilkorpora) enthalten.

Für die quantitative Auswertung wurden erstmals Verblisten mit den Verbkategorien mit Hilfe von *WMatrix* herausgefiltert, um die allgemeinen Verbhäufigkeiten in den Korpora feststellen, kategorisieren und vergleichen zu können – dabei lag der Fokus auf den POS-Tags VBG (*being*), VDG (*doing*), VHG (*having*) und VVG (alle weiteren lexikalischen Verben) für das *Progressive Aspect* und den POS-Tags VH0 (*have*), VHD (*had*) und VHZ (*has*) für das *Perfect Aspect*. Die sieben POS-Tags wurden dann manuell gefiltert und alle Formen, die nicht als *Progressive-* und *Perfect*-Formen fungieren (wie z.B. *going-to* Futur) wurden mit Hilfe von *WordSmith Tools* extrahiert und gelöscht. Darüber hinaus wurden die Häufigkeiten der *Progressive-* und *Perfect-* Formen auf der Basis von 1,000 Wörtern (*Tokens*) normalisiert. Als nächstes wurden die Häufigkeiten der finiten Verbalphrasen (Verbalphrasen kategorisiert mit den POS-Tags VBM, VBR, VBZ, VD0, VH0, VHZ, VV0 und VVZ) nach Halliday und James (1993: 39) und Quirk et al. (1974: 73) sorgfältig errechnet, um eine weitere Normalisierung der Häufigkeiten der Aspektformen vornehmen zu können – der angepasste

V-Koeffizient nach Smitterberg (2005), der die Häufigkeit der Aspektformen auf die Gesamtanzahl der finiten Verbalphrasen nach der folgenden Formel bezieht:

Frequency proportion aspect form (in %) = $\dfrac{\textbf{N aspect form occurrences} * \textbf{100}}{\textbf{N finite verb phrases}}$

Zusätzlich wurde die Kookurrenz der Aspektformen mit temporalen Adverbien und kontrahierten Hilfsverben quantitativ und funktional untersucht. Anschließend wurden die Lerner- und muttersprachlichen Häufigkeiten mit Hilfe von statistischen Tests (*Log-Likelihood Test*, vgl. Rayson und Garside 2000; Rayson 2008) verglichen, um Über- und Mindergebrauch feststellen zu können. Im nächsten Schritt wurden alle Aspektformen nach ihrem Vorkommen in Haupt- und Nebensätzen aufgeteilt und nach Vendlers (1957) vier Aktionsarten mit Hilfe von diagnostischen Tests (vgl. Brinton 1998: 242; Brinton 2000: 143 – 147) klassifiziert.

Im letzten Schritt wurde eine qualitative Auswertung des Lernergebrauchs von Aspektformen durch eine Muttersprachlerin (amerikanische Journalistin) vorgenommen, die als Ziel die Fehler-Markierung (*Error-Tagging*) aller grammatikalisch fehlerhaften Verbalphrasen hatte (die Muttersprachlerin war jedoch nicht über die Forschungsfrage informiert). Als Datengrundlage für das *Error-Tagging* dienten zwei Teilkorpora des *BUCLE* und *GICLE* mit ungefähr der Hälfte an Wörtern (ca. 110,000) – *BUCLE*_110,000 und *GICLE*_110,000 – die vollständig für Verbalfehler getaggt wurden. Darauf basierend wurden die Fehler analysiert und quantifiziert und als Grundlage für eine zielsprachliche Analyse herangezogen (*Targetlike Use Analysis*, vgl. Pica 1983), um die Fehlerhäufigkeiten der deutschen und bulgarischen Lerner anteilig an den Verbalphrasen vergleichen zu können und Rückschlüsse über den Einfluss der jeweiligen Muttersprache zu ermöglichen.

Kapitel 6. Quantitative Analyse

In diesem ersten Analysekapitel erfolgt zunächst eine kontrastive Bestandsaufnahme der finiten Verbalphrasen in den zwei Lernerkorpora *BUCLE* und *GICLE* und den vier muttersprachlichen Referenzkorpora *LOCNESS*_br, *LOCNESS*_us, *FLOB*_F und *FROWN*_F. Die Ergebnisse zeigen signifikante Unterschiede zwischen den Korpora, die auf einen Zusammenhang zwischen der relativen Häufigkeit der finiten Verbalformen und der Schreiberfahrung der Lerner und Muttersprachler hindeuten: je mehr Schreiberfahrung die Autoren der Texte in den Korpora haben, desto weniger finite Verbalphrasen verwenden sie.

Demzufolge bilden die Korpora die folgende Skala von verbalen bis hin zu nominalen Korpora:

**BUCLE > LOCNESS_us>GICLE>LOCNESS_br > FROWN_F >FLOB_F**

Diese Verteilung bestätigt die bisherigen Ergebnisse über die konzeptionelle Mündlichkeit in der Schriftsprache von fortgeschrittenen Lernern des Englischen (vgl. Altenberg 1997; Gilquin und Paquot 2007; Eriksson 2008 usw.) und spiegelt Koch und Oesterreichers (1985) Kontinuum von Mündlichkeit und Schriftlichkeit wieder: demnach sind *BUCLE* und *GICLE* „umgangssprachlicher" als die muttersprachlichen Schreibnovizenkorpora *LOCNESS*_br und *LOCNESS*_us, während die Expertenkorpora *FLOB*_F und *FROWN*_F eine deutlich höhere konzeptionelle Schriftlichkeit aufweisen.

Im nächsten Schritt werden die Häufigkeiten der *Progressive* Formen in den sechs Korpora anhand zweier Quantifizierungsmethoden ermittelt – relative Häufigkeiten von 1.000 Wörtern und relative Häufigkeiten bezogen auf die Anzahl aller finiten Verbalphrasen (*V*-Koeffizient nach Smitterberg 2005). Der kontrastive Vergleich beider Methoden zeigt den signifikanten Mindergebrauch in den Lernerkorpora *BUCLE* und *GICLE* – obwohl *GICLE* Lerner deutlich häufiger progressive Formen verwenden als *BUCLE* Lerner , liegen sie mit jeweils 2,45% und 1,7% progressiven Verbalphrasen immer noch unterhalb der meisten Werte in den muttersprachlichen Referenzkorpora (siehe Grafik). Die Unterschiede zwischen den muttersprachlichen Varietäten in Bezug auf die größere Präferenz für das *Progressive* in amerikanischem Englisch sind mit der Ausnahme von *FROWN_F* bestätigt. Die Mehrheit aller progressiven Verbalphrasen in den sechs Korpora ist für das Präsens markiert; jedoch weisen *GICLE* und die Expertenkorpora *FLOB_F* und *FROWN_F* mit ca. 25% *Past Progressive* Formen eine höhere Vergangenheitsorientierung auf.
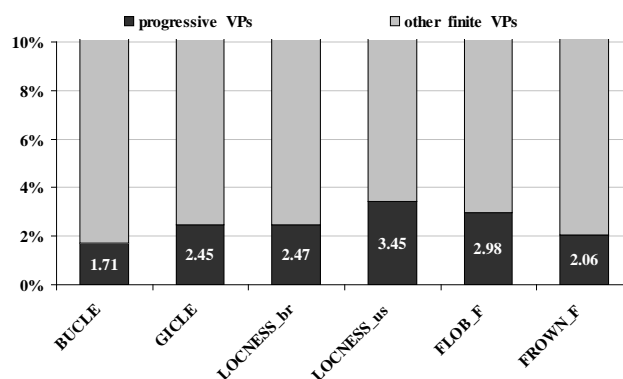
Anteil der progressiven Verbalphrasen in den sechs Korpora

Das gegenteilige Sprachverhalten weisen die Lernerkorpora *BUCLE* und *GICLE* in Bezug auf die relativen Häufigkeiten des *Perfect Aspects* auf: obwohl *BUCLE* Lerner deutlich häufiger Perfekt-Verbalphrasen als *GICLE* Lerner verwenden, liegen sie mit jeweils ca. 5,3% und 4,6% perfekten Verbalphrasen immer noch unterhalb der meisten Werte in den muttersprachlichen Referenzkorpora mit der Ausnahme von *FROWN_F* (siehe Grafik). Die Unterschiede zwischen den muttersprachlichen Varietäten in Bezug auf die größere Präferenz für das *Perfect* im britischen Englisch sind ebenfalls bestätigt. Die Mehrheit aller Perfekt-Verbalphrasen in den sechs Korpora ist für das Präsens markiert; ähnlich wie die temporale Markierung des *Progressive Aspect* weisen *GICLE* und die Expertenkorpora *FLOB_F* und *FROWN_F* mit ca. 28% *Past Perfect* Formen eine höhere Vergangenheitsorientierung auf.
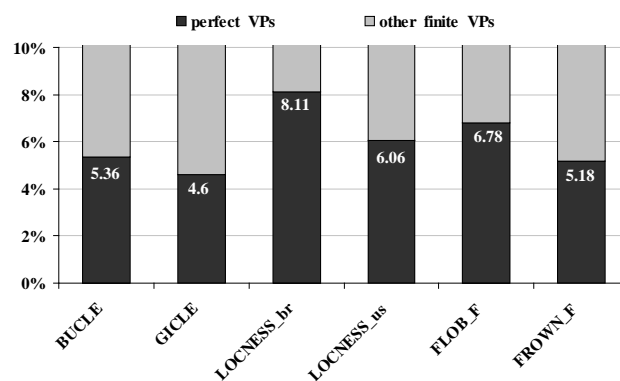

Anteil der perfekten Verbalphrasen in den sechs Korpora

Anhand des kontrastiven Vergleichs zwischen den relativen Häufigkeiten der beiden Aspektformen in den Lernerkorpora und den muttersprachlichen Referenzkorpora konnten Abweichungen von der muttersprachlichen korpusbasierten Norm bei beiden Lernergruppen verdeutlicht werden. Neben allen Unterschieden hat der Vergleich ebenfalls aufgezeigt, dass es auch gewisse Gemeinsamkeiten zwischen dem Gebrauch der bulgarischen Lerner mit dem in den britischen Korpora und dem Gebrauch der deutschen Lerner mit dem in den amerikanischen Korpora gibt.

Kapitel 7. Lexikogrammatikalische Analyse

Basierend auf den theoretischen Grundlagen der Aspekt- und Diskurshypothese, führt Kapitel 7 zunächst in die lexikogrammatikalische Analyse ein. Anschließend wird die Kookurrenz der Aspektformen mit Temporaladverbien und kontrahierten Hilfsverben untersucht. In Bezug auf die Aspekthypothese und den Zusammenhang zwischen Vendlers

Aktionsarten und Aspektmarkierungen in der Lernersprache lässt sich anhand einer *Across-Category* Analyse (vgl. Bardovi-Harlig 2000; 2002) feststellen, dass die Lerner in der vorliegenden Studie deutlich mehr telische Verben mit *Progressive*-Markierungen verwenden als die Muttersprachler. Gleichwohl liegt die Präferenz der bulgarischen und deutschen Lerner für atelische Verben mit progressiven Markierungen nicht nur an Tätigkeitsverben, sondern auch durchaus an statischen Verben, die sie im Vergleich zu den Muttersprachlern viel häufiger zusammen mit dem *Progressive* verwenden (siehe Grafik).
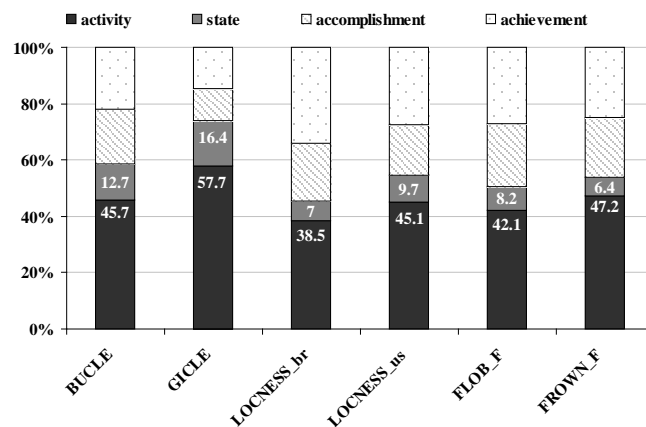
Verteilung der progressiven Verbalphrasen in Vendlers Aktionsarten

Dieses Ergebnis widerspricht der vierten Aussage der Aspekthypothese, dass Lerner progressive Markierungen zu statischen Verben nicht fehlerhaft übertragen. Besonders auffällig ist dabei die Präferenz der deutschen Lerner für statische Verben im *Progressive* – 7 von den 20 häufigsten Verben in dem *Progressive* in *GICLE* sind statische Verben, die nicht immer als zulässige Kombinationen im Sinne von Huddleston und Pullums (2005) Ausnahmen gelten.

Im Gegensatz zu den signifikanten Unterschieden zwischen den Lerner- und Referenzkorpora beim Gebrauch des *Progressive Aspect* sind keine so deutlichen Unterschiede beim Gebrauch des *Perfect Aspect* festzustellen: gleichwohl kommen telische Verben mit Perfekt-Markierungen in *GICLE* und *BUCLE* etwas häufiger vor als in den muttersprachlichen Korpora. Deutlich interessanter hingegen ist die *Type-Token* Relation (TTR) bei den *Progressive* und *Perfect* Formen in den sechs Korpora: die Lernerkorpora (mit der Ausnahme von *BUCLE* beim Progressive-Gebrauch) haben die niedrigste *Type-Token* Relation, während die muttersprachlichen Korpora (mit der Ausnahme von *LOCNESS*_br) ein Kontinuum von „niedrig" (bei den Schreibnovizenkorpora) bis „hoch" (bei den Expertenkorpora) bilden (siehe Grafik). Dieser Vergleich zeigt, dass bulgarische und deutsche Lerner eine recht begrenzte Anzahl von lexikalischen Verben mit dem *Progressive* und

Perfect verwenden – die häufigsten und „prototypischsten" Verben sind z.B. *be, live, sit, think, watch, become* usw. Die TTR-Unterschiede zwischen *BUCLE* und *GICLE* im Fall des *Progressive Aspect* liegen dabei nicht an einem „fortgeschritteneren" Wortschatz der bulgarischen Lerner, sondern an der Tatsache, dass bulgarische Lerner (ähnlich wie die Muttersprachler von *LOCNESS*_br) insgesamt sehr selten *Progressive Tokens* verwenden.

Type-Token Relation beim Progressive und Perfect in den sechs Korpora

Vor dem Hintergrund der Diskurshypothese werden im nächsten Schritt der Analyse die Häufigkeiten der *Progressive-* und *Perfect*-Verbalphrasen in Haupt- und Nebensätzen in den sechs Korpora miteinander verglichen. Beim Vergleich der *Progressive*-Verbalphrasen lassen sich keine auffälligen Unterschiede zwischen den Korpora feststellen: mit durchschnittlich ca. 55% der *Progressive*-Verbalphrasen in Hauptsätzen verwenden sowohl Lerner als auch Muttersprachler *Progressive*-Verbalphrasen meistens in Hauptsätzen (vgl. auch Biber et al. 1999: 461; Mindt 2000: 265); dennoch kommen *Progressive*-Verbalphrasen in den Schreibexpertenkorpora *FLOB*_F und *FROWN*_F etwas häufiger vor als in den Lerner- und Schreibnovizenkorpora. Demgegenüber ist die Verteilung der *Perfect*-Verbalphrasen nach Haupt- und Nebensätzen in den sechs Korpora erwähnenswerter: die bulgarischen und deutschen Lerner in der vorliegenden Studie verwenden mit durchschnittlich ca. 52% das *Perfect* deutlich seltener in Hauptsätzen als die Muttersprachler (siehe Grafik).
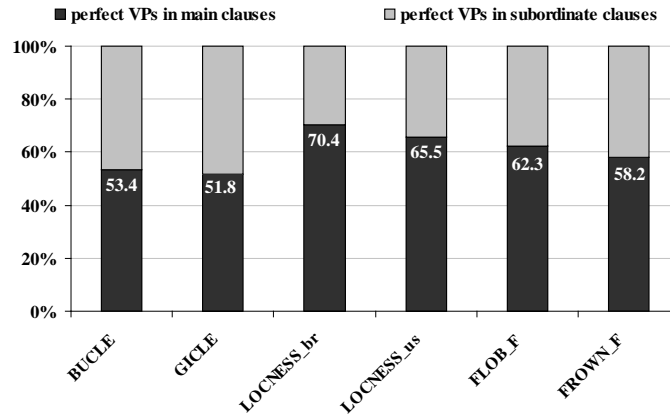
Häufigkeiten der Perfect-Verbalphrasen in den sechs Korpora

Diese Tendenz zum häufigeren Gebrauch des *Perfect* in Relativsätzen liegt unter Anderem daran, dass beide Lernergruppen das *Present* und *Past Perfect* wiederholt dafür verwenden, vorangegangene Ereignisse zu schildern, selbst wenn dieser Gebrauch fehlerhaft ist und die Zeitenfolge dadurch gestört wird, wie z.B. im folgenden Satz:

> 7.12. As **I have joined** the army for 12 months I have experience enough to assume that not only me but also about 90% of the boys whom I **have met** there would be willing and would be glad to make use of their "military knowledge" for "humanitary help" even if they run the risk to be shot. <*ICLE*-GE-AUG-0077.1>

Insgesamt zeigt sich, dass der Zusammenhang zwischen Erzählhintergrund und dem Gebrauch des *Perfect* stärker ist als der Zusammenhang zwischen Erzählhintergrund und dem Gebrauch des *Progressive*: die Lerner der vorliegenden Studie verbinden das *Perfect* häufiger mit dem Erzählhintergrund.

Im Anschluss an die lexikogrammatikalische Analyse wird die Kookurrenz der beiden Aspektformen mit Temporaladverbien und kontrahierten Hilfsverben untersucht. Die Ergebnisse über den Gebrauch des *Progressive* mit Temporaladverbien zeigen kein uniformes Bild; dennoch lässt sich feststellen, dass insbesondere bulgarische Lerner das *Progressive* vornehmlich zusammen mit Adverbien wie z.B. *constantly, still, nowadays und always* verwenden. Deutlich häufiger (über 30% im Durchschnitt) verwenden bulgarische und insbesondere deutsche Lerner das *Perfect* mit Temporaladverbien wie *always, never, just* usw.; dieses Sprachverhalten zeigt, dass bulgarische und deutsche Lerner des Englischen ihren Aspektgebrauch möglicherweise durch Temporaladverbien „verankern" möchten (vgl. Eriksson 2008; Granger 1998), um ihre Unsicherheit bezüglich des korrekten Gebrauchs der beiden Aspektformen kompensieren zu können.

Die Kookurrenz der beiden Aspektformen mit kontrahierten Hilfsverben in den sechs Korpora ist hier ebenfalls erwähnenswert, da es sich zeigt, dass deren Gebrauch Koch und

Oesterreichers (1985) Kontinuum von Mündlichkeit und Schriftlichkeit zum Teil wiederspiegelt: in Bezug auf die Verwendung von kontrahierten Hilfsverben mit dem *Progressive* und dem *Perfect* wie z.B. *I'm, you're* usw. sind *BUCLE* und *GICLE* umgangssprachlicher als die muttersprachlichen Schreibnovizenkorpora und *FLOB*_F und weisen somit eine deutlich höhere konzeptionelle Mündlichkeit auf. Die einzige Ausnahme stellt das amerikanische Schreibexpertenkorpus *FROWN*_F dar, das möglicherweise als die führende Varietät für den Sprachwandel im Sinne von „colloquialisation of written language" (vgl. Leech und Smith 2006: 199) fungiert (siehe Grafik).
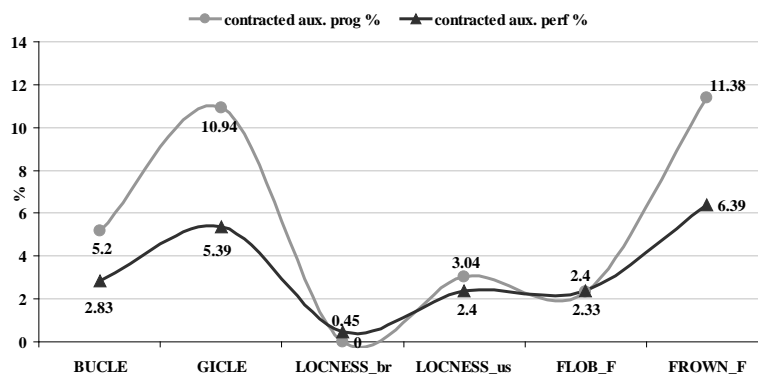
Anteil der kontrahierten Hilfsverben in *Progressive*- und *Perfect*-Verbalphrasen in den sechs Korpora

Kapitel 8. Fehleranalyse

Im Fokus des achten Kapitels steht die qualitative Analyse des Fehlgebrauchs der *Progressive*- und *Perfect*-Aspektformen in den Lernerkorpora *BUCLE* und *GICLE*. Das Kapitel führt zunächst in die neuesten Entwicklungen in der Fehleranalyse – die computergestützte Fehleranalyse (*Computer-Aided Error Analysis – CEA*, vgl. Dagneaux et al. 1998) – sowie die Methodologie der Fehleridentifizierung und Klassifizierung in der vorliegenden Untersuchung ein.

Die klassische Fehleranalyse (vgl. Corder 1967, Selinker 1972; 1976; Gass und Selinker 1994; Kellerman 1997, Sharwood Smith 1994, Ellis und Barkhuizen 2006 usw.) unterscheidet basierend auf der Fehlerursache zwischen interlingualen Fehlern – Interferenzen unter Einfluss der Muttersprache – und intralingualen Fehlern – Übergeneralisierungen in der Zweitsprache. Gleichwohl ist diese Unterscheidung nicht eindeutig, insbesondere wenn es sich um nicht-zielsprachlichen Gebrauch bei fortgeschrittenen Lernern bzw. fortgeschrittenen Stadien der Interimssprache handelt (vgl. Liu 2012; Kellermann 1997; Ellis und Barkhuizen 2006 usw.). Darüber hinaus existiert eine dritte Kategorie von Fehlern, die auf dem Sprachtransfer bedingt durch den Fremdsprachenunterricht basiert (*Transfer of Training*, vgl.

auch Selinker 1972; 1976; Gass und Selinker 1994). Für die fortgeschrittenen bulgarischen und deutschen Lerner in der vorliegenden Studie lässt sich vermuten, dass sie eher weniger interlinguale und mehr intralinguale sowie *Transfer-of-Training* Fehler machen werden; des Weiteren ist es von Bedeutung, dass die typologische und genetische Distanz zwischen den Muttersprachen Deutsch und Bulgarisch und der Zielsprache Englisch eine Rolle spielt.

Die Fehleranalyse in der vorliegenden Dissertation ist eine Erweiterung des computergestützten Fehleranalysemodells (CEA) nach Dagneaux et al. (1998), die eine qualitative und quantitative Untersuchung des Lernergebrauchs sowie Nicht-Gebrauchs der *Progressive-* und *Perfect-*Verbalphrasen beinhaltet. Dafür wurden alle von der Muttersprachlerin markierten *Error Tags* (siehe Kapitel 5) in den Teilkorpora *BUCLE*_110,000 und *GICLE*_110,000 ausgewertet, um Fälle der fehlerhaften Übergeneralisierung der beiden Aspektformen zu anderen temporalen Kontexten sowie Untergeneralisierung bzw. Nicht-Gebrauch in erforderlichen Kontexten feststellen zu können.

In Bezug auf die fehlerhafte Übertragung des progressiven Aspekts zu nicht-progressiven Kontexten lassen sich drei Haupttendenzen in beiden Lernerkorpora feststellen: 1) die Übertragung des progressiven Aspekts auf nicht-progressive Kontexte (z.B. generischen und allgemeinen Kontexten); 2) die Übertragung des progressiven Aspekts auf statische Verben und Verbalphrasen, und 3) der diskursbedingte Fehlgebrauch des progressiven Aspekts. In Bezug auf den Nicht-Gebrauch des progressiven Aspekts in erforderlichen progressiven Kontexten bzw. dessen Vermeidung (*Avoidance*) werden Unterschiede zwischen *BUCLE* und *GICLE* deutlich: in Situationen mit begrenzter Dauer (häufig auch in Zusammenhang mit Adverbien wie *now* und *nowadays*) meiden bulgarische Lerner den Gebrauch des progressiven Aspekts in einem höheren Maße als deutsche Lerner.

Hinsichtlich der fehlerhaften Übertragung des *Perfect* Aspekts zu anderen temporalen Kontexten lassen sich Gemeinsamkeiten zwischen den Lernerkorpora feststellen: beide Lernergruppen verwechseln die Funktionen des *Present Perfect* mit denen des *Simple Past* und verwenden das *Present Perfect* als deiktische Vergangenheitsform in narrativen Kontexten (auch in Kombination mit Temporaladverbien wie *yesterday* und *ago*). Demgegenüber ist der Nicht-Gebrauch des *Present Perfects* deutlich interessanter aus linguistischer Sicht, da insbesondere die deutschen Lerner das *Present Perfect* in erforderlichen *Present Perfect-*Kontexten eher meiden.

In der anschließenden zielsprachlichen Analyse nach Pica (1983) wird versucht, die Unterschiede zwischen den Fehlgebrauchsquoten in den Lernerkorpora quantitativ festzuhalten. Dafür werden die Häufigkeiten aller finiten Verbalphrasen sowie die

Häufigkeiten der progressiven und perfekten Verbalphrasen in den Teilkorpora *BUCLE*_110,000 und *GICLE*_110,000 neu ermittelt. Auf dieser Grundlage wird der Fehlgebrauch der beiden Aspektformen gemessen und miteinander verglichen, indem die Häufigkeiten der fehlerhaften Übertragung sowie die des Nicht-Gebrauchs der beiden Aspektformen auf die ermittelte Anzahl deren erforderlichen Kontexte bezogen wird (siehe Grafik).
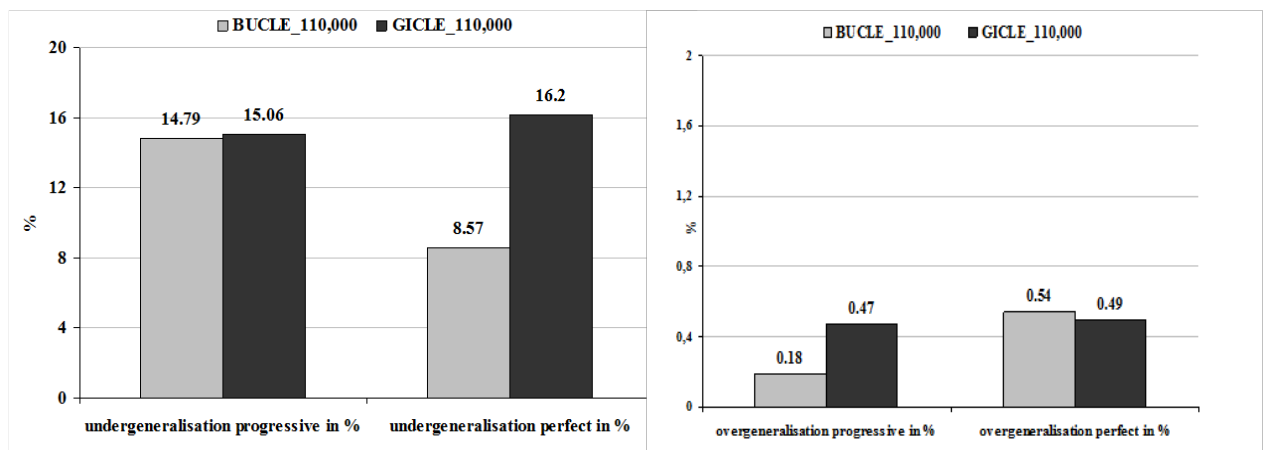
Über- und Untergeneralisierung des Progressive- und Perfect-Aspekts

Die zielsprachliche Analyse zeigt, dass der Nicht-Gebrauch der Aspektformen in erforderlichen Kontexten für beide Lernergruppen deutlich problematischer ist als der fehlerhafte Übergebrauch. Des Weiteren lassen sich Unterschiede zwischen den beiden Lernerkorpora feststellen: so wird das *Perfect Aspect* von den deutschen Lernern häufiger gemieden, während das *Progressive Aspect* sich als besonders schwierig für die bulgarischen Lerner erweist. Insgesamt ist es auffällig, dass die relativen Häufigkeiten der zielsprachlich verwendeten progressiven und perfekten Verbalphrasen in den beiden Lernerkorpora auf einen noch höheren Mindergebrauch bzw. Nicht-Gebrauch gegenüber der muttersprachlichen korpusbasierten Norm hindeuten.

Kapitel 9. Diskussion und Modellbildung

Die aus der quantitativen und qualitativen Analyse gewonnenen Erkenntnisse werden in diesem Kapitel zusammengeführt und evaluiert und im Anschluss wird ein Modell zur korpusbasierten Analyse des Lernergebrauchs von Aspektformen erstellt. Vor dem Hintergrund der quantitativen Ergebnisse wird die Verteilung der Tempus- und Aspektformen in den sechs Korpora erneut beleuchtet und miteinander verglichen: dabei fällt auf, dass beide Lernergruppen überdurchschnittlich viele Präsensformen verwenden, die möglicherweise anstelle von anderen Tempus- und Aspektformen wie z.B. dem *Present Progressive* oder dem

Present Perfect fehlerhaft verwendet wurden, um eine täuschende Tempuskontinuität zu erzeugen. Ein mögliche Erklärung dafür ist der unvollständige Erwerb der Formen und Funktionen der englischen Tempus- und Aspektformen (und insbesondere komplexerer Tempus- und Aspektformen) bei fortgeschrittenen bulgarischen und deutschen Lerner des Englischen und der daraus resultierende Übergebrauch der „sichersten", unmarkierten, prototypischsten Form – der *Simple Present* Form. Zudem ist es denkbar, dass sich die Lernerkorpora und die muttersprachlichen Referenzkorpora gattungstechnisch unterscheiden, da insbesondere die Schreibexpertenkorpora *FLOB*_F und *FROWN*_F eine deutlich höhere innere Variabilität auf lexikalischer, aber auch grammatikalischer Ebene aufzeigen.

Aus den Ergebnissen der lexikogrammatikalischen Analyse lässt sich schließen, dass beide Lernergruppen sich von den inhärenten semantischen Eigenschaften der Verbalphrasen beim Gebrauch des *Progressive Aspect* beeinflussen lassen, d.h. beide bevorzugen atelische Verben mit progressiven Markierungen. Des Weiteren übertragen beide Lernergruppen das *Progressive* auf statische Verben und Verbalphrasen (z.B. kognitive Verben) und verwenden das *Perfect Aspect* vornehmlich in Nebensätzen. Die falsche Übertragung des *Progressive Aspect* auf statische Verben entgegen den Aussagen der Aspekthypothese liegt unter Anderem daran, dass beide Lernergruppen insgesamt eine sehr kleine Anzahl der meist gebrauchten Verben (sowohl statisch als auch dynamisch) im Englischen verwenden, und sich innerhalb von diesem kleinen Wortschatz von den Aktionsarten beeinflussen lassen. Entgegen der Diskurshypothese präferieren beide Lernergruppen und insbesondere *GICLE* Lerner *Perfect*-Verbalphrasen mit telischen Verben vornehmlich in Nebensätzen bzw. im Erzählhintergrund; dies liegt auch daran, dass die Lernertexte eher auf Erzählungen und Beschreibungen basieren, die das *Present* und *Past Perfect* häufig fehlerhaft als Vorvergangenheitsformen verwenden.

Die übergeordnete Leitfrage beim kontrastiven Vergleich zwischen dem Gebrauch der Aspektformen durch fortgeschrittene bulgarische und deutsche Lerner des Englischen bezieht sich darauf, inwiefern es Unterschiede zwischen den beiden Lernergruppen gibt und inwiefern diese Unterschiede auf Interferenzen aus der Muttersprache zurückzuführen sind. Eine erneute Untersuchung der gemiedenen progressiven Formen im Teilkorpus *BUCLE_110,000* und deren äquivalenten Übersetzungen in der Muttersprache L1 Bulgarisch nach dem *Integrated Contrastive Model* von Gilquin und Granger (2008; 2002) zeigt, dass in 100% der Nicht-Gebrauch-Fällen die entsprechende bulgarische Form das imperfektive Verb beinhaltet. Somit

lässt sich der starke Mindergebrauch der progressiven Formen in *BUCLE* mit dem Einfluss der Muttersprache gut erklären – bulgarische Lerner des Englischen sehen „keinen Bedarf" für progressive Markierungen, da sie sie mit dem bulgarischen imperfektiven Aspekt verwechseln, das synthetisch anstatt analytisch flektiert wird. Demgegenüber haben deutsche Lerner eher Schwierigkeiten mit dem *Perfect Aspect*: die gleiche Analyse der entsprechenden Übersetzungen der gemiedenen *Present Perfect* -Formen in *GICLE* im L1 Deutsch zeigt, dass deutsche Lerner das *Present Perfect* konzeptionell als reine Vergangenheitsform betrachten und die *Present Perfect*-Formen daher häufig durch *Simple Past*-Formen ersetzen, jedoch auch (wenn auch seltener) in *Simple Past*-Kontexten verwenden. Eine konzeptionelle Übertragung der Funktionen des bulgarischen Perfekts wird ebenfalls deutlich beim Betrachten des Übergebrauchs des *Present Perfect* in *BUCLE* und der entsprechenden Übersetzungen im L1 Bulgarisch: bulgarische Lerner betrachten das englische *Perfect* unter Anderem als indirekte Erzählform und verwenden es, um Wissen „zweiter Hand" auszudrücken.

Der Einfluss der Muttersprache kann für den nicht-zielsprachlichen Gebrauch von Aspektformen nicht als einziger Grund herangezogen werden; weitere außersprachliche Parameter wie das Sprachkompetenzniveau, die Schreiberfahrung der Lerner, der Einfluss der länderspezifischen Fremdsprachenvermittlung (z.B. *Transfer of Training*) und nicht zuletzt der Kontakt zu englischsprachigen Muttersprachlern spielen ebenfalls eine Rolle. Die quantitative Analyse hat verdeutlicht, dass die Lerner- und muttersprachlichen Varietäten ein Kontinuum in Bezug auf konzeptionelle Schriftlichkeit bilden; darüber hinaus lassen sich Gemeinsamkeiten zwischen *BUCLE* und der britischen korpusbasierten Norm sowie *GICLE* und der amerikanischen korpusbasierten Norm in Bezug auf den Gebrauch von Aspektformen feststellen. Diese werden folgendermaßen abgebildet:
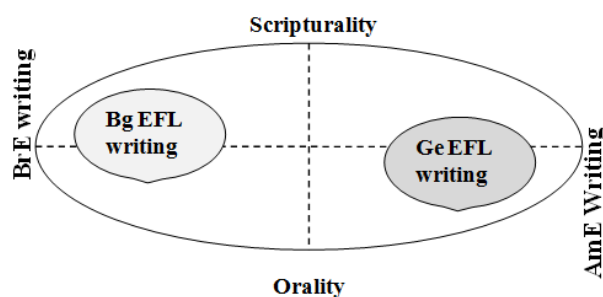

Zweidimensionale Darstellung von dem bulgarischen und deutschen Schriftenglisch

Zusätzlich ist es von Bedeutung, dass im Gegensatz zu den einigen wenigen bulgarischen Lernern, über 50% der deutschen Lerner mindestens einen Monat im

englischsprachigen Ausland verbracht haben; dazu wurden sie als fortgeschrittener eingestuft. Demzufolge ist es denkbar, dass deutsche Lerner möglicherweise Kontakt zu amerikanischen Muttersprachlern hatten, was deren Sprachverhalten erklären könnte. Der Einfluss der länderspezifischen Fremdsprachenvermittlung wie z.B. die „Überbetonung" der Formen und Funktionen des *Perfect Aspect* im deutschen EFL Klassenzimmer oder die allgemeine Fehldarstellung der beiden Aspektformen in deutschen und bulgarischen EFL Büchern und Lernmaterialien können ebenfalls für den nicht-zielsprachlichen Gebrauch verantwortlich sein.

Im letzten Teil dieses Kapitels wird basierend auf den Erkenntnissen ein integriertes Modell zur Untersuchung von Aspektformen in fortgeschrittenen Lernervarietäten vorgeschlagen. Das Modell ist ein Prozess, der vier konsekutive Schritte beinhaltet: 1) die kontrastive Häufigkeitsanalyse der Aspektformen in der Lernersprache und Zielsprache Englisch, 2) die lexikogrammatikalische und kontextuelle Analyse der Aspektformen, 3) die computergestützte kontrastive Fehlerhäufigkeitsanalyse, und 4) die zielsprachliche Analyse. Dieses Modell hat den Vorteil, dass es die quantitativen korpuslinguistischen Methoden mit den qualitativen Methoden der Zweitspracherwerbsforschung verbindet und damit die Lücken der klassischen *Contrastive Interlanguage Analysis* schließt.
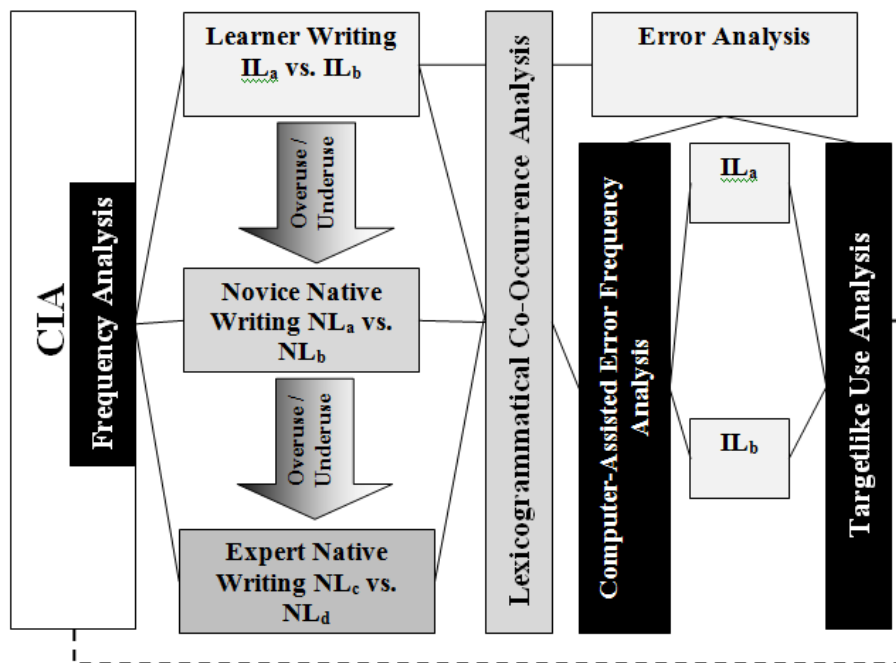
Integriertes Modell zur Untersuchung von Aspektgebrauch in der Lernersprache

Zusammenfassend lässt sich sagen, dass eine Vielzahl an Faktoren für den nicht-zielsprachlichen Gebrauch des *Progressive* und *Perfect Aspect* durch fortgeschrittene deutsche und bulgarische Lerner des Englischen verantwortlich ist, wie z.B. der

muttersprachliche Einfluss, der Einfluss der länderspezifischen Fremdsprachenvermittlung, das Zusammenspiel zwischen Lexik und Grammatik, der Kontakt zur Zielsprache sowie Lern- bzw. Vereinfachungsstrategien.

Kapitel 10. Schlussbemerkung und Ausblick

Das letzte Kapitel fasst die Ergebnisse zusammen und gibt einen Ausblick auf mögliche Entwicklungen der Lernerkorpusforschung an der Schnittstelle zur Zweitspracherwerbsforschung. Vor diesem Hintergrund regt die Arbeit eine stärkere Zusammenarbeit zwischen experimentellen Forschern und Korpuslinguisten an; zum Anderen wird für eine stärkere Berücksichtigung von allen Einflussfaktoren (sprachlich sowie außersprachlich) auf den Lernervarietäten appelliert. Weitere interessante Fragen beziehen sich auf die Rolle der expliziten vs. impliziten Zielnorm im Zweitspracherwerb bzw. Fremdsprachenunterricht sowie die Rolle der Verflechtung von Tempus, Aspekt und Modus im Zweitspracherwerb. Schlussfolgernd hat die Untersuchung fremdsprachendidaktische Implikationen in Bezug auf die Notwendigkeit der Vermittlung des Aspekts im Kontext und der stärkeren Fokussierung auf Schreibtechniken in Fremdsprachenunterricht.

Erklärung zur Urheberschaft

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Aus fremden Quellen Übernommenes ist kenntlich gemacht.

_____, den _____ _____
(Ort) *(Datum)* *(Unterschrift des Verfassers)*