5 Bamberger Beiträge zur Linguistik

The Present Perfect in World Englishes

Charting Unity and Diversity

von Valentin Werner

University of Bamberg Press

5 Bamberger Beiträge zur Linguistik

Bamberger Beiträge zur Linguistik

Band 5

hrsg. von Manfred Krug, Thomas Becker, Martin Haase,
Sebastian Kempgen und Patrizia Noel Aziz Hanna

University
of Bamberg
Press
2014

The Present Perfect in World Englishes

Charting Unity and Diversity

von Valentin Werner

University
of Bamberg
Press
2014

Bibliographische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.ddb.de/ abrufbar

Diese Arbeit hat der Fakultät Geistes- und Kulturwissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.
1. Gutachter: Prof. Dr. Manfred G. Krug
2. Gutachter: PD Dr. Julia Schlüter
Tag der mündlichen Prüfung: 13. November 2013

Herstellung und Druck: Druckerei docupoint, Magdeburg
Umschlaggestaltung: University of Bamberg Press, Andra Brandhofer

ISSN: 2190-3298
ISBN: 978-3-86309-225-2 (Druckausgabe)
eISBN: 978-3-86309-226-9 (Online-Ausgabe)
URN: urn:nbn:de:bvb:473-opus-68834

Acknowledgements

This book grew out of a dissertation project at the Chair for English and Historical Linguistics at the University of Bamberg, while parts of the research were carried out at the Department of Theoretical and Applied Linguistics of the University of Cambridge. I would like to thank my supervisor, Manfred G. Krug, for having made all this possible in the first place, for inciting my interest in the intricacies of empirical language study, and for providing detailed feedback while leaving me time and room to develop my ideas. I would also like to extend my gratitude to many people and institutions for contributing in various ways.

Above all, the thesis profited from input from the co-supervisors, Julia Schlüter and Geoffrey Haig. Both supported the project from its early stages and provided constructive criticism throughout. The same counts for Mari C. Jones, who acted as advisor during my research stay in Cambridge.

Michaela Hilbert and Ole Schützler ventured to comment on earlier written chapters of the draft and special mention needs to be made of Alison Edwards, who did a marvelous proofreading job.

Subsections of the thesis were presented and discussed at various conferences. Thanks are due to the audiences and in particular to Thomas Brunner, Bernd Kortmann, Raj Mesthrie, Joybrato Mukherjee, Devyani Sharma, Cristina Suárez-Gómez, Peter Uhrig, and Jim Walker for valuable comments and suggestions.

Travel to these meetings was made possible by generous financial support from the German Academic Exchange Service (DAAD), the International Society for the Linguistics of English (ISLE), St. Catharine's College Cambridge, the University of Bamberg, and the Volkswagen Foundation.

As regards the more technical, but no less important side, I would like to thank all ICE teams for their efforts in corpus compilation and the staff of the university libraries in Bamberg and Cambridge for always being ready to help with a smile. Advice on statistics, data processing, and visualization, given at different stages by Robert Fuchs, Martin Hilpert, Ghada Mohamed, and Lukas Sönning, was also highly appreciated.

In addition, I would like to thank my colleagues at the Chair for English and Historical Linguistics and the people of the Bamberg Graduate School of Linguistics (BaGL) for providing a friendly and stimulating working environment, and for being uncomplicated and knowledgeable people with whom to share thoughts and ideas.

My parents and friends, and particularly Carolin provided extensive non-academic support throughout my studies. It is to them that this work is dedicated.

List of abbreviations

AAE	Afro-American English
AusE	Australian English
BNC	British National Corpus
BrE	British English
BROWN	Brown University Standard Corpus of Present-Day American English
CanE	Canadian English
COCA	Corpus of Contemporary American English
COHA	Corpus of Historical American English
EAfE	East African English
EFL	English as a Foreign Language
EIL	English as an International Language
EModE	Early Modern English
ENL	English as a Native Language
ESL	English as a Second Language
FijiE	Fiji English
F-LOB	Freiburg-LOB Corpus of British English
FROWN	Freiburg-Brown Corpus of American English
HKE	Hong Kong English
ICE	International Corpus of English
ICLE	International Corpus of Learner English
IndE	Indian English
IrE	Irish English
JamE	Jamaican English
L1	First Language
L2	Second Language
LCSAE	Longman Corpus of Spoken American English
LOB	Lancaster-Oslo-Bergen Corpus
ME	Middle English
MOP	Medial object perfect
NigE	Nigerian English
NZE	New Zealand English
OE	Old English
PDE	Present Day English
PhiE	Philippine English
pttw	per ten thousand words
SinE	Singapore English

List of figures

List of tables

1 Introduction

According to Bauer, the Present Perfect (HAVE + V-*en*; henceforth PrPf) "has always been a somewhat inconvenient case" (1970: 189) in research on English grammar. The elusiveness of the form can be deduced from statements such as that by Labov below, which highlights the notorious problems in describing its semantics:

> The meaning of *I have lived here*, as opposed to *I lived here*, has been explained as (a) relevant to the present, (b) past *in* the present, (c) perfective, (d) indefinite, (e) causative, and so on. It is not only that there are many views; it is that in any given discussion no linguist has really found a method by which he can reasonably hope to persuade others that he is right. (1978: 13)

Apparently, not much has changed since the 1970s, despite persistent additions to the body of works on the topic. Klein observes that "[t]here is a very rich literature on the English perfect [...], but one cannot say that there is a generally accepted analysis" (2009a: 54). In a similar vein, Veloudis states that "although it has been one of the most discussed problems in the theory of grammatical categories, the perfect eludes a convincing analysis" (2003: 385). Others propose that such a uniform analysis cannot be achieved for the English PrPf at all due to its multifaceted nature (McGilvray 1991: 47). This broad functional range inherent in the form makes it a testing ground *par excellence* for studying variation in the expression of *temporality*[1] as one of the "essential component[s] of verb meaning" (Klein 1994: xi). The present study focuses on both internal variation (i.e. variation along language-internal factors) and external variation (i.e. variation along language-external categorizations).

In contrast to the ample extant theoretical literature, corpus-based studies of the PrPf in varieties of English are comparatively scarce. As

[1] A note on notation: capital letters are used for crosslinguistic category types (e.g. PERFECT), small capital letters for language-specific category types (e.g. ASPECT), initial capitals for language-specific categories (e.g. Present Perfect) and italicization for notional semantic categories (e.g. *time*) whenever necessary in order to discriminate different layers of description.

yet, the vast majority of these studies have restricted their focus to differences between BrE and AmE as the two "traditional" (i.e. most widely studied/arguably with the longest linguistic history) varieties of English, have merely considered variation between the PrPf and the Simple Past (henceforth SPst) or other alternative surface forms, or have considered the PrPf a "learner-hard" feature. Therefore, the present empirical investigation is an attempt to fill a gap in the description of PrPf usage in a number of further varieties. It takes a synchronic, contrastive perspective and uses a corpus-based approach. The present work demonstrates the very vitality of the paradigm of World Englishes, which (i) has become a well-established area of linguistic research and continues to expand, for instance in terms of the growing amount of comparable corpus material within the scope of the ICE project, and (ii) contributes to a sounder understanding of the structures and usage patterns of the English language in general. Based on the analysis of the corpus findings, the study tackles the following issues:

- Can the PrPf be seen as a core feature of the grammar(s) of different World Englishes or (and to which extent) does nativization emerge? Along which dimensions (e.g. geographical location, variety types, and text types) can we observe variation? Are there any particular epicenters that determine distributional patterns?
- Is it possible to establish register and genre effects within and across varieties, applying categories from Biber's analytical framework (Biber 1988; Biber & Conrad 2009)?
- Do the data contain non-standard uses of the PrPf or alternative perfect forms? If so, how do these forms pattern? Does this have any ramifications for the grammatical status of the PrPf?
- What is the central (both language-internal and language-external) context in which the PrPf occurs in World Englishes?

The study also considers the particular role that temporal adverbials play in the variation between the SPst and the PrPf. In addition, it discusses the adequacy of current models of World Englishes for embedding the

findings with respect to individual structural features (as exemplified in the present study) in a larger theoretical context.

This study begins with an outline of the theoretical and historical foundations (Chapter 2) that is necessary to establish a general framework into which the study of the English PrPf can be integrated. It sketches issues such as the dichotomies *time* and *temporality*, *tense* and *aspect*, *perfect* and *perfective* as well as the typological and historical location of the English PrPf. This is followed by a brief introduction to the study of World Englishes and related influential models and concepts. Subsequently, Chapter 3 contains a detailed review of previous research in order to systematize the many accounts of the PrPf and to discuss their implications for the present investigation. Chapter 4 is devoted to a description of the methodology and the data that served as the groundwork for the findings. It comprises descriptions of the structure of the data, of the procedures involved in tagging, identifying and annotating the variables, and of the multidimensional aggregational methods used.

Chapters 5 and 6 present the results of the empirical analysis. Chapter 5 considers ten varieties individually (AusE, CanE, BrE, HKE, IndE, IrE, JamE, NZE, PhiE, SinE), highlighting the distributions of the individual language-internal and influences of the language-external variables. Chapter 6 then provides a bird's eye perspective of the data. This part abstracts from many individual observations in order to identify global trends and to assess the convergence of the varieties under scrutiny, again taking account of language-external categorizations and testing the applicability of theoretical notions such as *areoversals* or *varioversals*. The chapter concludes with three case studies, (i) on surface forms that can be used as alternatives to the PrPf and their semantic and pragmatic properties, (ii) on the combinability of temporal adverbials with the PrPf and the SPst, and (iii) on evidence in ICE for usage of the PrPf as a tense that claims territory from the SPst. In Chapter 7, the main points emerging from the corpus study are condensed and discussed. At the same time, this chapter revisits issues raised at the outset of the study in light of the data presented. In this way, the corpus findings are further related to the research context as well as to broader theoretical issues. Finally, Chapter 8 sums up the main findings and con-

tains some reflections on methodological aspects. In addition, it sketches potential avenues worth further investigation.

2 Theoretical and historical foundations

2.1 Time, temporality and typological aspects

As a first step toward obtaining a picture of the Present Perfect in English, this section will explore the relationship between *time* and *tense* in general and the typology of the PERFECT in particular. This is necessary for a comprehensive understanding of the theoretical background, which underpins the analyses of theories of the semantics and pragmatics of the Present Perfect as well as the corpus-based investigations presented later (chapters 3, 5 and 6). In particular, the input from language philosophy and language typology is considered.

The concept of *time*[2] has received a great deal of attention in scholarly disciplines as diverse as physics (e.g. as part of discussions of the theory of relativity), philosophy (e.g. in the context of phenomenological studies on the relationship between space and time) and philology (e.g. within different narrative theories and modes of narration). Within this wider framework, the importance of conveying time-related messages is an elementary feature of human communication around the world and can therefore be seen as a true anthropologic fundamental (Comrie 1985: 7; Klein 2009a: 35).

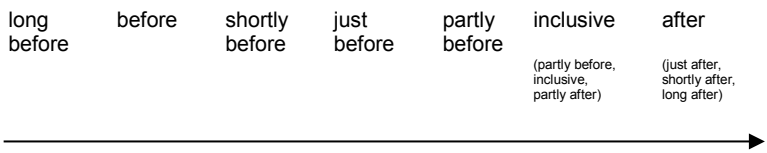


Figure 2.1.1. Temporal relations according to Klein (2009a)

In general, *time* is seen as an independent extralinguistic entity whereas *tense* is a purely linguistic category[3] (Declerck 2006: 94), and these two

[2] I will not attempt a universally valid definition of *time*, as this may be impossible (Klein 2009a: 6) and in any event of only minor interest to the present study, as this section is concerned with the conceptualization of *time* in language. An overview of various approaches toward *time* in the course of human history can be found in Klein (2009a).

[3] I follow Klein & Vater in their definition of grammatical category as "mapping between particular formal means and particular meanings (or functions)" (1998: 215). Their caveat,

levels should not be confused. Yet, because the conceptualization of *time* in different cultures is by no means unidimensional, Klein (2009a: 35–36) has cautioned that uniform accounts of both *time* and its conceptualization in language are inclined to oversimplification. In addition, he identifies a number of temporal relations that can potentially be expressed via *tense* (see Figure 2.1.1), but, with due credit for his attempt to include as many potential temporal relations that may be expressed in languages as possible, his list may be criticized for being redundant (e.g. "partly before" appearing twice in the model) and not particularly explicit about transitions from one label to another (e.g. criteria for distinguishing between "before" and "long before" and the hierarchical relationships between concepts; Klein 2009a: 31–32).

Next, it is helpful to consider the ways in which languages can convey *temporality*; after all, *tense* is the most prominent, but not the only option that is available. The list below provides an overview of the principal means of encoding *time* and temporal relations in language (based on Klein 1994: 14 and Klein 2009b: 40–41). Not all of the following are directly relevant to the subsequent analysis of English, but they are mentioned here for the sake of completeness:

- Tense (e.g. the English Simple Past/Preterit)
- Aspect (e.g. the English Progressive)
- Temporal adverbials (e.g. adverbs)
- Aktionsart/inherent temporal features of the verb (e.g. punctuality, durativity)
- Temporal particles (e.g. perfectivity markers in Asian languages)
- Principles of discourse organization (e.g. reported order corresponds to chronological temporal order in reality)

The above list could be further complemented by secondary devices, for example complex verb clusters such as *to continue to smoke* or *to begin to sleep* (Klein 1994: 14), or time-bound nominals with inherent temporality

---

that "unity of formal marking and of specific meaning [...] is rarely found in human language" (Klein & Vater 1998: 215) is especially true in the case of the English PrPf.

that is expressed lexically such as *ex-mayor* or *the future Prime Minister* (Binnick 1991: 448). In the following, I will discuss the main options in more detail, expanding on the general characteristics of *tense*, *aspect*, *Aktionsart* and temporal adverbials.

The expression of *time* is obligatory in almost all natural languages (see below) and more often than not achieved via *tense*, usually involving a structural connection to the finite verb (Declerck 2006: 94; Vater 2007: 42).[4] The view taken in many older studies is that "tense indicates whether the situation described by a sentence is in the past, present, or in the future – in other words, whether it precedes, is simultaneous to, or follows the speech event" (Klein 2009b: 51; see the approach taken by Pulgram 1987 for an example). This tripartite definition at first glance seems elegant and straightforward, and indeed it has been considered conventional wisdom for centuries (Jaszczolt 2009: 82). It has come under criticism, however, as (i) *tense* is not the sole option to express *temporality* (see the list above), (ii) although the vast majority of languages have ways to express *tense* (Comrie 1985: 9), languages without *tense* in the classical sense (i.e. with grammatical marking) do in fact exist (e.g. Burmese; Comrie 1985: 50–53; cf. Dahl & Velupillai 2011d), (iii) this definition may be too abstract to incorporate the manifold temporal relations that actually exist (Klein 1992: 536), and (iv) it does not include "non-canonical usages" (Klein 2009b: 48) such as narrative presents describing past events.

In his widely accepted definition, Comrie states that *tense* is "the grammaticalized expression of location in time" and that "much of what has traditionally been called tense falls under that category" (1985: vii). More recent, but substantively similar versions are provided by Jaszczolt, who proposes to "draw the boundary between tense and time in a more intuitive way [by saying] that tense is the grammatical means of conveying temporality of the situation" (2009: 84); Klein, who states that *tense* "serves to hook up the topmost verbal element of the clause-internal temporal structure to the clause-external temporal structure" (2009b: 77–78); and Declerck, who builds a bridge to linguistic form in stating that

[4] However, others see this structural view as too narrow, and claim rather that propositions are tensed (Vasudeva 1985: 188).

tense "is the pairing of a morpho-syntactic form with meaning, the meaning being the specification of the temporal location of a situation" (2006: 94). It has been argued, however, that *tense* is not a sufficient device even to encode *temporality*, as languages with *tense* often complement it with one or more of the further options, temporal adverbials in particular (Jaszczolt 2009: 82). Nevertheless, although "its role for the expression of time may be a bit overrated in the research tradition" (Klein 2009b: 51), *tense* is commonly understood as the most important means of expressing *temporality*, at least in Indo-European languages. At the same time, languages tend to have tense systems that are much more fine-grained in referring to the past than in referring to the future (Comrie 1985: 85–101).

Grammatical *aspect*, on the other hand, can often not be distinguished from *tense* in its function of expressing *temporality* (Comrie 1976: 66; Dahl & Velupillai 2011a).[5] Suffice it to say here that it is usually concerned with the internal temporal structure (Comrie 1976: 3). Similar to *tense*, but on a different level (see Figure 2.1.2), it is connected to the "time of the top-most verbal element" (Klein 2009b: 78).

On yet another level, but comparable to and interacting with grammatical aspect in relation to the temporal structure within a clause, *Aktionsart* (sometimes also referred to as "lexical aspect", "intrinsic aspectual character" or "inherent aspect") can be defined as "the modes in which an action or activity, and generally an event, process or state proceeds" (Bartsch 1995: 128). It refers to the "inherent time-argument structures" (Klein 2009b: 78) of verbal expressions, which are ascribed values along dimensions such as qualitative change (stative vs. non-stative), boundaries (telic vs. atelic), duration (punctual vs. non-punctual), inner quantification (iterative vs. semelfactive) or phase (in-

[5] Cf. the alternative approach presented in Dahl & Velupillai (2011a): "An alternative to seeing tense, aspect and mood as grammatical categories in the traditional sense is to regard tense-aspect-mood systems as wholes where the building-blocks are the individual tenses, aspects, moods, such as the Past and the Progressive in English." This approach is based on the finding that it is fairly uncommon to encounter languages that possess either *tense* or *aspect* instead of both or neither (Dahl & Velupillai 2011b). As a detailed discussion of the English PrPf with regard to its *tense* and *aspect* properties and the disputed notion of *current relevance* that underlies most typological accounts will follow in Section 3.2, no further comments are added at this stage.

choative, terminative, continuative) (Klein 1994: 79–80; for alternative categorizations see e.g. Meyer 1992). A more detailed account of the model used in the present study will be provided in Section 4.3.2 below.

A further option used to express temporal location in many languages is that of temporal adverbials. Unlike the options mentioned above, it is not part of grammar. The crucial difference between temporal adverbials and the grammaticalized devices is that "[w]hile temporal adverbials are used only when they express information which is relevant to the particular intended message, the use of tenses [and aspects; V.W.] is guided by general principles that often make the choice of a certain tense [or aspect; V.W.] obligatory and that make the use of a tense morpheme obligatory even if the information it carries is redundant" (Dahl & Velupillai 2011a). Again, as more specific information can be found below (Section 4.3.1), suffice it to say at this point that temporal adverbials have been categorized according to their form and function within the scope of a variety of approaches (e.g. Quirk et al. 1985; Panzner 1995; Panitz 1998, to name but a few with respect to English).

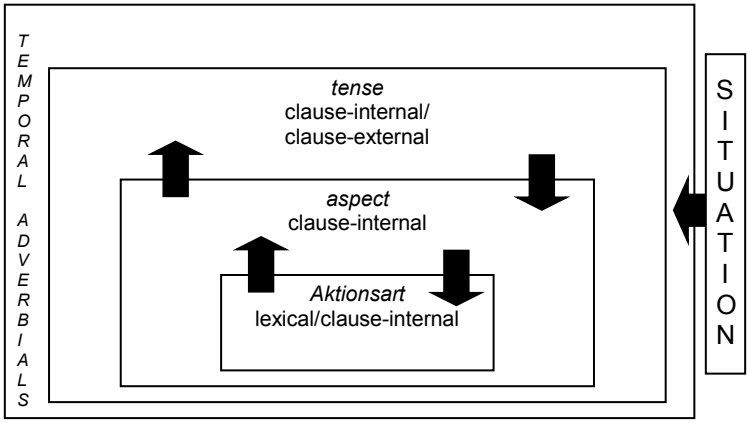


Figure 2.1.2. Ways of expressing *temporality* in languages

Figure 2.1.2 presents a simplified summary overview[6] of the most salient ways of expressing *temporality* in languages, considering the various levels of interaction between the individual levels.

Following this short introduction to the ways in which temporal relations can be expressed in languages in general, I now proceed to some of the typological aspects of the PERFECT. After a few remarks on the history and development of the PERFECT worldwide, I discuss the spread and types of PERFECTS, and finally address the English PrPf. At first sight, it may not be clear why such information is needed for a synchronic account of varieties of Present Day English (PDE). However, it is undertaken in order to provide a wider linguistic backdrop for the findings of the present study, and to determine whether developments in other languages influence the current patterns of usage in these varieties. Independent of the specific languages involved, typological tendencies also give insight into cognitive aspects, such as *processability* (Pienemann & Keßler 2012), and feature in many language contact hypotheses that take account of issues such as *markedness* (Bybee 2011; cf. Eckman 1977; see Section 2.5.1 below).

The history of the PERFECT has received a considerable amount of scholarly attention in the field of language typology and diverse contrastive linguistic studies (among others Bybee et al. 1994; Dahl 1995; Dahl & Velupillai 2011c). A number of potential PERFECT sources (sometimes also referred to as "anteriors", to use Bybee et al.'s terminology) have been identified (after Dahl 1995: 19; cf. Bybee & Dahl 1989: 57; Bybee et al. 1994: 87–91):

- Possessive constructions (e.g. *I have the house built*)

[6] This model could be criticized as an oversimplification, especially in light of comprehensive studies on *temporality* in language such as Klein (1994) or Thompson (2005). However, it is not intended to present an alternative generic model of the encoding of *time* in language, as the intricacies of such a model would go far beyond the scope of the present study. Moreover, there is no universally agreed model of *temporality*. Rather, Figure 2.1.2 and the preceding explanations suggest one way of how the different devices that are available can be used to express *temporality*. From a performative perspective, "situation" could be defined as "the particular time line invoked in the mental representation of a history – a representation which the speaker presumes is shared by the hearer" (Michaelis 1994: 121).

- Participles or converbs in predicative position with or without a copula
- Constructions involving a verb with the general meaning 'finish'
- Constructions involving morphemes meaning 'already'
- Movement verbs

Interestingly, whatever the source of the PERFECT, a common path of development along the cline RESULTATIVE[7] > PERFECT > PERFECTIVE[8] > PAST can be observed diachronically in a number of dissimilar and (geographically and genetically) unrelated languages (Bybee & Dahl 1989: 56, 68–77; Schwenter 1994: 996; Boogart 1999: 134; Engel & Ritz 2000: 124). In many languages other than English, perfects have reached one of the last two stages. This entails both a generalization of their meaning and a loss of other meaning components:

> [T]he change of an anterior to a past or perfective is typical of grammaticization changes. On the semantic level, the change is clearly a generalization of meaning, or the loss of a specific component of meaning: the anterior signals a past action that is relevant to the current moment, while the past and perfective signal only a past action. The specification of current relevance is lost. The meaning generalizes in the sense that the past or perfective gram expresses a more general meaning that is compatible with more contexts. [...] The anterior conveys the sense of past or perfective but includes a special flavor of relevance or proximity to the present or current situation. Thus if a speaker wishes to frame his or her contribution AS THOUGH it were highly relevant to current concerns, then the

[7] Bybee et al. define the resultative as "a present state [...] described as the result of past action" and provide several secondary readings of perfects, namely (i) "experiential, in which certain qualities or knowledge are attributable to the agent due to past experiences" and (ii) "anterior continuing, in which a past action continues into present time" (Bybee et al. 1994: 62; for a contrastive analysis of resultative vs. perfect see also Bybee & Dahl 1989: 68–70). Another recognized reading is the expression of recentness (Dahl & Velupillai 2011c). In contrast, "grammaticalized marking of time in the form of tenses is typically independent of considerations of relevance" (Dahl & Velupillai 2011d).

[8] Note that "perfect" and "perfective" are not the same thing, although different authors (wrongly) use the terms interchangeably. In typological accounts, "perfect" implies merely anteriority and current relevance *sensu* Bybee et al. (1994: 54), while "perfective" implies a discrete action that is completed or finished (see also Comrie 1976). Section 2.3 provides a more detailed discussion.

> speaker might use the anterior more often than would strictly be necessary for the communication of the propositional content of the message. Such overuse weakens the force of the current relevance component, and eventually the hearer infers only past or perfective action from the anterior and no sense of current relevance. (emphasis original; Bybee et al. 1994: 86; see also 105 (Figure 3.1))

This generalization has led to the present situation evident especially in European languages: What used to be PERFECT has evolved into a PAST that is in competition with (older) PAST forms, as is the case in Spanish or German, or has ousted them already, as in French (Engel & Ritz 2000: 122–124; Hantson 2005: 251; Ritz & Engel 2008: 138; Schaden 2009: 138), potentially due to the early influence from classical Greek via Parisian French and subsequent swift expansion during the Middle Ages (Drinka 2003: 11; Dahl & Velupillai 2011c).

With regard to the current spread of perfect types, observations from large-scale typological studies such as GRAMCATS (Dahl 1995) and WALS (Dryer & Haspelmath 2011), which compare grammatical descriptions of an extensive sample of languages, provide some insight.

Table 2.1.1. Distribution of perfects in languages worldwide

Perfect type	**GRAMCATS**[9]		**WALS**	
HAVE-perfect	2	(2.6%)	7	(3.2%)
Perfect with 'finish' or 'already'	6	(7.9%)	21	(9.5%)
Other (e.g. BE-perfect)	31	(40.8%)	80	(36.0%)
No perfect	37	(48.7%)	114	(51.3%)
Total number of languages in the sample	*76*	*(100%)*	*222*	*(100%)*

[9] In detail, GRAMCATS found eleven perfects with BE, three with 'finish', three with 'already', four involving movement verbs, and sixteen labeled "other" (Dahl 1995: 20).

What can be inferred from Table 2.1.1 is that (i) HAVE-perfects are rather infrequent globally and (ii) at least half of the world's languages do not have the grammatical category PERFECT at all, which suggests that it represents a typologically marked category *sensu* Eckman (1977). Furthermore, the two studies suggest that perfects are restricted geographically, being found mostly in Europe and South and Southeast Asia,[10] while 'already'/'finish' perfects are prominent mostly in Southeast Asia and West Africa.[11] More precisely, HAVE- and BE-perfects are predominantly found in Europe (as constructions involving copula verbs are frequent in this area; Dahl 1995: 20) with the contemporary function of general pasts (see above). Some authors (e.g. Drinka 2003: 3–6) claim an east-west split with HAVE-perfects in Western Europe (with exceptions such as German or Italian) and BE-perfects in Eastern and Central Europe. On a related note, it remains to be tested whether the occurrence of 'already'/'finish' perfects carries implications for the usage of the PrPf in regional varieties of English in areas where this type is salient in the substrate language(s).

As regards form and function, in addition to the points given above, two general statements about perfects in languages worldwide are relevant in relation to the English PrPf. First, with regard to form, the English PrPf seems rather prototypical, as perfects often involve periphrastic (copula/auxiliary + past participle) constructions (Dahl 1985: 129; Bybee & Dahl 1989: 56; Krug 2011; Dahl & Velupillai 2011c).[12] Second, with regard to function, Bybee et al. claim that perfects often occur in "conversational discourse" (1994: 62) and signal current relevance

[10] Thus, within current typological versal theory, perfects of the European type seem to represent typical "areoversals" (Szmrecsanyi & Kortmann 2009: 33).

[11] The prominence of the latter type could be explained by the fact that morphological marking of tense and aspect is largely absent in general in these areas (Dahl & Velupillai 2011c) and the expression of *perfect* by non-morphological, in this case lexical, means functionally compensates for this potential communicative need.

[12] Note that pasts (and perfectives) are usually marked inflectionally, as they are viewed as being more advanced on the grammaticalization path (see above), where inflectional marking is much more likely to occur (Dahl & Velupillai 2011a; 2011d).

(see above), while suggesting that perfects are "typically translated with the English Perfect" (1994: 54).[13]

Aside from the findings that serve to identify the English PrPf as a prototypical perfect from a diachronic perspective, there are a number of points that establish its special status from a synchronic typological point of view, in particular when taking the European context into consideration. As has been pointed out repeatedly, one of the unique features of the English PrPf is that, unlike in many other European languages, its resultative meaning is persistent (Miller 2000: 343).[14] This persistence can be further deduced from a number of properties, such as indistinctive occurrence with all verb forms, unification of its formal outlay (HAVE + V-*en*), or (quasi-)mandatory use in certain temporal environments (Pietsch 2009: 537). In addition, currently there seems to be no indication that the status of the English PrPf[15] is likely to evolve further along the suggested grammaticalization path toward PERFECTIVE or PAST. This seems to go against the general trend in other European languages, where, on the whole, perfects are characterized by their instability (Ritz & Engel 2008: 137) and periphrastic perfects have claimed the territory of morphological pasts, as described above. It should be added that the English PrPf is also unique in its non-compatibility with certain temporal adverbials (Mugler 1988: 62–63).[16] Speculations on the

[13] They recognize, however, that it may as well appear as a narrative tense in some cases (Bybee et al. 1994: 62).

[14] A possible reason why the English perfect has not grammaticalized further is provided by Drinka (2003: 21–22), who speculates that the development from PERFECT > PAST only occurred in the core area of Europe. This would explain why perfects in other arguably peripheral languages, such as Castilian, Catalan, or Swedish, have also retained the "anterior" meaning.

[15] At least in what are considered "old standards" of English, such as AmE and especially BrE. The situation is different for other varieties of English, as will be shown repeatedly in this study.

[16] This statement has to be taken with a grain of salt, as combinations of the PrPf with temporal adverbials that would normally be considered ungrammatical or odd do in fact occur in restricted grammatical contexts (e.g. PrPf infinitives of the type *Peter is thought to have called her yesterday* or afterthoughts; see Anderson 1982: 230). The same is true for a particular group of temporal adverbials (such as *recently* or *just*) referring to the recent past (Comrie 1985: 84; see further Section 6.5 below). The hypothesis that combinations are possible in other languages due to special features of the present tense in these languages (Portner 2003: 497) cannot be fully rejected, as indeed it seems possible for the present

theoretical feasibility of the English PrPf developing into a true PERFECTIVE (Pulgram 1987: 390) or PAST (Sempere-Martinez 2008: 135) and anecdotal evidence for the unrestricted combinability of the English PrPf with temporal adverbials (Miller 2000: 350) are repeatedly raised based on observations of the development of other European languages. For the time being, however, corpus-based studies of general corpora (e.g. Hundt & Smith 2009) have revealed no substantial changes within this grammatical area, at least in standard varieties.

Therefore, the typologically special status of the English PrPf[17] seems even more surprising, as languages that are genetically and areally close are usually prone to undergo parallel or at least similar developments (Sempere-Martinez 2008: 135) and as a result to have similar structural features (Muysken 2010: 265). In other words, with regard to form and function in the wider (mainly diachronic) typological context, the English PrPf represents a relatively typical PERFECT (Bybee et al. 1994: 61),[18] but from a narrower (mainly synchronic and areally limited) contrastive view, the English PrPf is unique in its persistence as a grammatical category, being more marked (i.e. functionally more restricted) than perfects in other European languages, which have been generalized to establishing past reference (Schaden 2009: 130).[19]

---

tense to take over functional domains of the perfect (e.g. German *Er wohnt seit 20 Jahren in diesem Haus*). However, this does not provide an account of why the situation is different in English, and is thus only of limited explanatory power.

[17] As other languages have constructions that are formally similar but not functionally congruent (Siemund 2004: 419), this also requires specific awareness and attention in translation. The continuing importance of the topic of the PrPf within translation studies and in the context of language teaching is underscored by the vast number of contrastive investigations that exist (e.g. Lucko 1981 for German; Pulgram 1987 for Romance and German; Schwenter 1994 and Downing 1996 for Spanish; Molsing 2006 for Portuguese; Rothstein 2008 and 2011 for German and Swedish; Boogart 1999 and Landman 2008 for Dutch).

[18] The claim that perfects are "often accompanied by the relational adverbs 'already' and 'just'" (Bybee et al. 1994: 54), however, only applies to the English PrPf with strong restrictions. Quantitatively it is not fully verifiable, as will be shown below (see chapters 5 and 6). Cf. also Kortmann (1995: 195), who lists further counter-arguments to the view that the English PrPf is a typical PERFECT.

[19] These languages (e.g. German and French) have seen a complementary development where typologically "old" pasts have now become the marked option that is restricted e.g. to certain registers (Schaden 2009: 126).

2.2 Tense and aspect in English

Having briefly considered *time*, *temporality* and the PERFECT from a typological perspective, including the special situation of the English PrPf in comparison to other perfects in Europe and worldwide, I now proceed to discuss TENSE and ASPECT specifically in English. This section explores the general grammatical framework into which the PrPf is embedded; a detailed analysis of which grammatical category (or categories) the PrPf belongs to will follow in Chapter 3.

First of all, there is no clear agreement on the number of TENSES in English (König 1995: 153–156), with counts ranging between 2 and 36. It is evident that these counts depend very much on the scope of the definition applied, for example whether periphrastic structures (such as the *will*-future) are included or not. Descriptive linguistic accounts are normally found at the more conservative end of the spectrum, as they either apply formal criteria (e.g. Jespersen 1931: 3, who posits two TENSES, the "Present" and "Preterit")[20] or logical criteria (e.g. Meyer 1995: 202, who establishes TENSE as a category parallel to PROGRESSIVE and PERFECT, as all three have binary values – such as "present" or "past" for TENSE – for each verbal element). Some authors cover the middle ground; they either see the *will*-future as a properly grammaticalized TENSE in English (Anderwald 2009: 49), or do not posit ASPECT as a separate category and thus include more forms within their TENSE schemes (for a pertinent example see Declerck 2006, who claims that "verb forms are tense forms if they relate a situation time directly or indirectly to the zero-point" (22), and therefore includes the PrPf but excludes modal expressions).[21] Lastly, learner grammars and usage guides usually tend

[20] He acknowledges that "Tense-Phrases" (Jespersen 1931: 3) as subordinate categories exist, including the PrPf, which "is itself a kind of present tense, and serves to connect the present time with the past" (Jespersen 1931: 47; for more recent accounts in the same vein cf. Ota 1963: 41 and Salkie 1989: 8). All of these authors, however, remain vague when it comes to establishing the exact nature of this connection (but cf. Ritz & Engel 2008: 150 for an attempt within Klein's temporal framework).

[21] Cf. also the notions of *absolute* and *relative tense* (König 1995: 157), which are based on the approach first introduced in Reichenbach (1947) and were further elaborated in Comrie (1985).

to be most inclusive in their presentation of the English tense system and may often incorporate MODALITY, VOICE and periphrastic forms into their descriptions, which results in an extended number of TENSES.[22]

The number of ASPECTS in English is also disputed, but most authors agree that there is at least one, namely the "progressive" form in English. The discourse tends to focus instead on the membership of "perfect" in the ASPECT category, and the majority of accounts recognize two aspects, namely the "progressive" and "perfect", which can also be combined.[23] The former is formally indicated by BE + V-*ing* and signals that a situation is continuous and non-stative or temporary, with other potential features, such as contingency (Comrie 1976: 32–40). The latter is formally marked by HAVE + V-*en* and is highly versatile in its semantic and pragmatic implications, as will be shown in the discussion below. The downside of this "rather eccentric behavior" (Schaden 2009: 116) is that it is impossible to find a universally valid description, which is further complicated by the fact that a specific category such as the English PrPf can be even more variable within the context of an empirical study based on data from different speaker communities and registers (Ritz & Engel 2008: 139).

Clearly, anchoring the PrPf within the TENSE-ASPECT framework of English is no trivial task. The study of the multiple janiform nature of the PrPf – that is (i) that it cannot be easily categorized into the two grammatical categories described above, (ii) that it can refer to both present and past (Boogart 1999: 133, 150), and (iii) that it is interchangeable with the SPst in some contexts (Mittwoch 2008: 346)[24] – has resulted in a flood of publications, but not yet in a commonly accepted theory. All this, however, makes it an intriguing linguistic testing ground.

---

[22] See Swan (2006) or www.ego4u.com/en/cram-up/grammar/tenses for examples. A Google query on "tenses English" returns various versions of tense tables with variable numbers of tenses for learners. Functional grammar also recognizes an extended number of tenses, 36 in the case of Halliday & Matthiessen (2004: 340–342).

[23] Note that "imperfective" and "progressive" are not necessarily the same thing (Comrie 1976: 33–34; but cf. Klein 1992: 538).

[24] This can depend on the pragmatic conceptualization of the individual speaker (Declerck 2006: 147; see further sections 3.3.3, 6.6 and 7.2). Cf. also Dahl & Hedin's (2000: 387) theory of type-focusing ("Asking for how many" → PrPf) and token-focusing ("Asking whether there are any" → SPst).

2.3 Perfect and perfective

So far, precisely what constitutes a PERFECT, a PERFECTIVE and an English PrPf has simply been tacitly assumed. As we are now heading toward an analysis of the English PrPf as such, a few notes on terminology are in order.

As pointed out above, typological accounts normally employ the label PERFECTIVE if a construction implies a discrete action that is completed or finished (Dahl & Velupillai 2011a), and PERFECT to indicate mere *anteriority* and *current relevance* (see also Dahl 1985: 138–139; Klein 1994: 110).[25] This has led to confusion particularly in linguistic studies of English, as (i) a characteristic trait of the English PrPf in contrast to perfects in many other languages is that, depending on the different semantic interpretations of the construction, it is not necessarily PERFECTIVE at all (Inoue 1979: 561; Elsness 1997: 18; Nordlander 1998: 5),[26] and (ii) with reference to the terminological issue mentioned above, Quirk et al. (1985) employ the term PERFECTIVE when referring to PERFECT in the "anterior" sense.[27] In spite of repeated criticism (e.g. by Kortmann 1991: 16–17 and Nordlander 1998: 3), many other scholars have simply reproduced Quirk et al.'s terminological oddity without any further evaluation or use different labels interchangeably (see Meyer 1995: 204; Fritz 2006: 290, 291; Landman 2008: 131; Bongartz & Buschfeld 2011: 43 for pertinent examples).[28]

[25] Other features that are typically associated with PERFECTIVE, such as short duration or resultativeness, are provided in Binnick (1991: 154). There is also an alternative perspective suggesting that PERFECT merely expresses posteriority to a process (Mugler 1988: 235; Klein 1994: 109–110).

[26] Cf. also Portner (2003: 466), who maintains that the English PrPf is PERFECTIVE but that this is not its defining characteristic, as other TENSE-ASPECT entities (such as the SPst) share PERFECTIVITY as a feature. This claim is partly acceptable but weakened by the fact that combinations of PERFECT and IMPERFECTIVE (realized by progressive ASPECT) are possible (e.g. *He has been mowing the garden for two hours now*).

[27] They are, however, aware that "[t]raditionally the term PERFECT has been frequently used instead of PERFECTIVE" (Quirk et al. 1985: 3.54)

[28] See also www.ucl.ac.uk/internet-grammar/verbs/tense.htm.

The current account seeks to avoid these pitfalls, sticking to the most common label "(English) PrPf"[29] (in contrast to the typological aspectual categories PERFECT and PERFECTIVE). As pointed out above, the object of investigation is typically formed using HAVE + V-*en*, although there are other formal variants with semantic values similar to the PrPf. Due to the nature of the data used in the present study, a formal approach toward identifying instances of the PrPf is primarily employed, but other constructions with the semantic value of the PrPf are identified and described where applicable (see further Section 4.2).[30]

2.4 On the history of the English Present Perfect

Although the current study is synchronic in nature, a few points on the history of this form in English specifically are in order, both to comprehend the wider context of the PrPf and to reconstruct the developmental path that led to the current special status of the PrPf. This should also enable us to follow several considerations that lie behind previous accounts of the subject.

As pointed out in the section on typological aspects above, perfects may have different sources. For English, principally one possibility is posited, although historical linguists agree that an adequate reconstruction of the exact source and the subsequent development is only possible by approximation (Wright 1986: 113; Michaelis 1998: 122–124). According to most accounts, the original source of the PrPf in English are possessive-resultatives of the type *I have the house built* ('I possess the house in a built state'), which underwent an extension with regard to their semantic implications (from RESULTATIVE to ANTERIOR) and

[29] For a list of more exotic labels for the PrPf used since 1781, see Wynne (2000: 8).

[30] With this I follow the approach suggested by Meyer (1995: 201–204), who indicates that purely semantic definitions of the PrPf fall short as they are based on notions of traditional Latin-derived grammar. Further drawbacks compared to a formal approach are that (i) semantic definitions are more prone to subjectivity, and (ii) from a pragmatic point of view, occurrences of constructions with a PrPf semantic value are difficult to identify automatically within corpus data and thus require extensive manual analysis. A semantic definition cannot be avoided, however, if non-standard forms expressing a meaning similar to the PrPf should be included in the analysis (see e.g. Section 6.4 below; cf. Davydova 2011).

finally grammaticalized into PERFECT (Garcia 1985: 277–285; Schwenter 1994: 1006–1011; Bybee et al. 1994: 68–69; Veloudis 2003: 394; Miller 2004a: 230; Michaelis 2006: 223; McFadden & Alexiadou 2010: 392–394; see also Askedal 2010 and Section 2.1 above),[31] while HAVE lost its "possess" implication (Rainer 1989: 1). This development could have stemmed from a re-analysis within the construction HAVE (*habban*) + (NP) object + passive participle (*I have the wall painted*)[32] of the last constituent as a past participle (Hantson 2005: 248). Others maintain that even initially the participle had "a good deal of adjectival force" (Michaelis 1998: 123; cf. Wynne 2000: 11). Both interpretations agree, however, that the element that used to modify the object by referring to its state (Boogart 1999: 135) was re-interpreted, and as a result an inference of previous completed action ultimately led to the emergence of the PERFECT interpretation.

While older accounts acknowledge the existence of a periphrastic perfect in English from the ME period onwards at best (Åkerlund 1911: 85), its development is claimed to have already begun at the beginning of the OE period (e.g. McFadden & Alexiadou 2010: 392) or even earlier, in proto-Germanic (Wynne 2000: 11).[33] Even if the exact details of the timeline are far from clear, it seems likely that the initial spread of the English PrPf was analogous to the spread of perfects in other European

[31] An interesting finding that would merit a study in its own right is that the path of acquisition of the PrPf by children seems to follow the historical path of development (first solely "resultative", then later extension to further semantic facets such as "continuative" or "experiential"), although different processes are involved (Slobin 1994: 129; cf. Diessel 2011 for discussion). Further, it would be worth investigating whether such a path of acquisition (or a path with a different order) can also be established for speakers of non-native varieties of English.

[32] Constructions of this type still occur (e.g. as causatives; Hantson 2005: 248) and are alleged to be especially prominent in IrE (see sections 5.2 and 6.4 below), possibly due to the fact that the Irish Perfect had a similar source (Pietsch 2009: 531–536). Note that BE-perfects following a similar structural development are not considered here as they have never been as frequent as HAVE-perfects (Jespersen 1931: 29–46). See Wright (1986: 113–165) and McFadden & Alexiadou (2010: 394–406) for detailed accounts of the recession and functional restriction of BE-perfects.

[33] Cf. Brinton (1988: 99–102), who claims that PERFECT was an established category in Germanic, as combinations of HAVE (+ object) + past participle with possessive meaning were already rare in OE. She argues that the positional change of the elements was an indication of semantic change rather than its cause.

languages, which can be situated "at the time of transition between antiquity and the Early Middle Ages" (Drinka 2003: 2). Nevertheless, during the transition from OE to ME and further in ME, the position of the object evidently became more flexible and more and more intransitive verbs could be found in the position of the last constituent. Explanations for this greater flexibility include (i) the weakened congruence between object and participle (Rainer 1989: 1–4) and (ii) the process of "exbraciation", that is, the main verb/participle moved next to HAVE while in turn the object and adverbials were postponed (Wynne 2000: 12). A number of contexts that favor the PrPf have been identified, among them particular groups of verbs (mental state/perception/reporting and telic verbs in general), while other contexts (such as atelic verbs, anterior adverbials, duratives, statives and iteratives) are seen as favoring PrPfs at a later stage only (Carey 1994, 1995; cited in Van Herk 2010: 53).

In relation to the functions of the PrPf in contrast to the SPst, while the principal means for past time reference in OE was the SPst (Pulgram 1987: 389), PrPf forms became increasingly common. From the ME period continuing into EModE, there was a certain amount of variation between PrPf and SPst forms (Michaelis 1998: 122–124; Tagliamonte 2000: 330; but cf. Vermant 1983: 3), with the PrPf even occasionally serving as a narrative tense (Wright 1986: 135).[34] In certain genres such as letters, however, the regulation of and functional division between the PrPf and SPst seems to have been at a fairly advanced stage from the beginning of the EModE period onwards, so that the vast majority of occurrences of PrPf forms were in line with PDE rules (Rainer 1989: 4). However, these findings may need qualification, as the overall frequency of PrPf constructions had not yet reached PDE levels (Wynne 2000: 15) and its functional distribution in other genres, especially in spoken language, was probably less established.

Yet, even a conservative estimate would see the underlying rules of distribution in PDE as having been in operation from the end of the EModE period. This does not mean that in PDE there is no variation;

[34] Some older studies provide an alternative explanation, with the PrPf historically moving along the continuum TENSE > ASPECT. This view has largely been discarded in the majority of diachronic investigations (Vermant 1983: 3).

after all, both formal and functional variants of the PrPf have existed and do exist. In particular, scholarly attention has been attracted by the differences between AmE and BrE usage (e.g. Meyer 1995; Elsness 1997, 2009b, 2012; Wynne 2000; Schlüter 2002a; Hundt & Smith 2009) and by variation in other geographical, social and learner varieties of English (e.g. Tagliamonte 2000; Payre-Ficout et al. 2009). As additional points will be discussed in the review of previous theoretical and empirical studies (see Chapter 3 below), I will not go into further detail about the status of the PrPf in PDE at this stage. Suffice it to add here that the functional range (Van Herk 2010: 57) and the frequency of the PrPf have constantly increased in the history of English (presumably taking over ground from the SPst, as argued by McCoard 1978: 149 and Van Herk 2008: 49; see further McFadden & Alexiadou 2010: 415), but that only minor changes could be observed in the last century (Hundt & Smith 2009: 48, 57).[35] Therefore, although PERFECT may be considered a volatile category from a diachronic and typological point of view, the status of the English PrPf is in fact very stable synchronically and any substantial change – such as a universal development into a narrative TENSE (Pulgram 1987: 390–391; Sempere-Martinez 2008: 135; Schaden 2009: 138) or, on the contrary, the ousting of the PrPf by the SPst (a scenario discussed by Ahlqvist 2000: 171) – is considered unlikely within the intermediate future (Vermant 1983: 109). Still, it cannot fully be excluded that PrPf usage patterns as present in older stages of English or that developments relating to early language contact during colonization have persisted in some varieties (see further sections 2.5.1 and 6.4 on the "founder principle"; Mufwene 1996; cited in Hickey 2004).

[35] Therefore, Hundt & Smith argue that a "narrowing of the gap" (2009: 48) between AmE and BrE as claimed by Elsness (2009b) is implausible. Note, however, that Elsness (2012) provides quantitative evidence from a range of diachronic and synchronic corpora (including COHA and COCA) that PrPf frequencies have indeed decreased since the 18th century in both AmE and (less markedly) in BrE, although some of his explanations for this development remain tentative.

2.5 The study of World Englishes

Another area of research relevant to the present study is that of World Englishes, a branch of linguistics that is particularly associated with the postcolonial era and that has increased in popularity since the 1960s.[36] For the present purposes, this section discusses three aspects. First, it considers the various models of World Englishes, with a view to determining whether the distinctions between varieties that are drawn in the various models are also reflected in the usage of the PrPf, or whether the PrPf rather represents a "core" feature of English[37] with little internal variation. Second, it assesses the strength of the forces of cultural globalization on language; here the underlying hypothesis is that cultural globalization leads to globalization of language, in other words to relatively uniform usage independent of variety type but potentially (still) oriented toward one or more of the "old" varieties, such as BrE and particularly AmE. Finally, it explores the relatively new subfield of the study of "versals" within the World Englishes context in an effort to test the applicability of one or more of the versal concepts for one particular area of grammar.

[36] For a discussion of the politics and polemics that have surrounded the study of World Englishes, see Minnich (2004: 17–23), who argues that "'English' is a cultural and constructed concept, rather than a natural phenomenon" (2004: 15). She thus suggests the alternative label "worldly englishes" (with intentionally lowercase letters) in an effort to avoid any political bias.

[37] Note that some authors have stated that World Englishes are an excellent testing ground as "[i]n this field, variation can be studied among different varieties of one language, which still share a common lexical core" (Wolf & Polzenhagen 2009: xii). Others have noticed that tense marking, for example, is relatively consistent, i.e. mostly corresponding to standard patterns, across all varieties of English (Mesthrie & Bhatt 2008: 58). It remains to be tested whether other facets of grammar, and in particular PrPf usage, should also constitute such a "core" or "internationalized" (Hundt et al. 2004: 588) area.

2.5.1 Models

This section addresses a number of models that seek to describe the situation of English in various locations,[38] mostly from a sociocultural perspective. An earlier metastudy of models of World Englishes (Minnich 2004) identified 20 models in total (not including later additions such as Schneider's (2007) model or the model employed by Szmrecsanyi & Kortmann (2011)). For reasons of space, only the most influential of these models are presented here.

It has been noted that models of World Englishes can be categorized according to either their formal layout (e.g. circles and family trees) or their theoretical approach (e.g. spatial, historical, evolutionary and phaseal), all of which carry metaphorical implications. Another common distinction is between "static" models (e.g. the circle, family tree and spatial approaches) and "dynamic" models that take into account sociocultural developmental patterns (e.g. phase models). It can be argued that the former are more oriented toward the past, while the latter also have some form of predictive potential (Minnich 2004: 127–142). It is also worth noting that the majority of the models are composed of units that are geographically and politically defined (Minnich 2004: 199), which emphasizes the importance assigned to spatial relationships in the context of World Englishes.

One of the oldest models is that proposed by Strevens (1992: 33). His derivational family tree model (see Figure 2.5.1) is based on the underlying assumptions (i) that the fundamental distinction is that between BrE and AmE, and (ii) that all other varieties can eventually trace their "ancestry" to these two varieties. Thus, this model mainly reflects colonial history and possibly also geopolitical realities (as argued by Saraceni 2010: 61), suggesting that all native and non-native varieties on a lower level of the hierarchy are still oriented toward either one or the other of the donor varieties (e.g. CanE and PhiE should be closer to

[38] Cf. Mesthrie & Bhatt (2008: 27–36), Jenkins (2009: 17–24), and Saraceni (2010: 61–69) for contrasting views of models of World Englishes.

AmE, NZE and IndE should be closer to BrE, etc.). The model makes no claims about the current status[39] of the varieties.

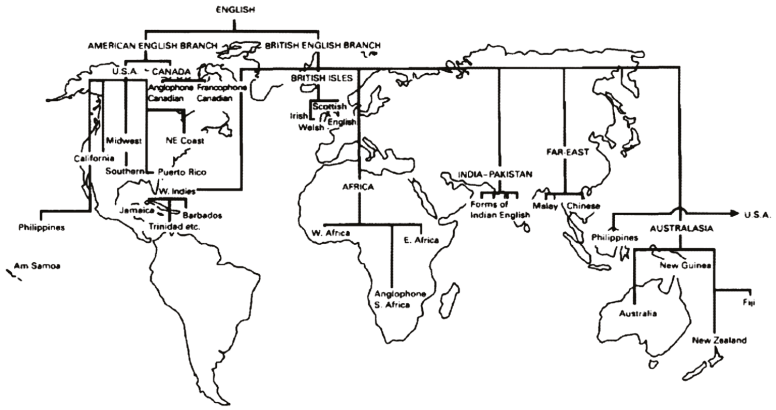


Figure 2.5.1. Strevens's (1980/1992) model of World Englishes

The next models to be outlined all conceptualize and categorize World Englishes based on a circle metaphor. While they are apparently similar at first glance, they nevertheless contain some subtle yet important differences.

The most influential of the circle models is that developed by Kachru (1988). He establishes three circles of English (see Figure 2.5.2), whose individual members are political entities that are assigned to the different circles based on historical and functional principles. In detail, the "Inner Circle" comprises varieties that resulted from settlement colonization[40] (besides BrE as the source) and that are assigned a "native" status. Therefore, these varieties are seen as "first" or "native language varieties" (L1/ENL) and as "norm-providing" for others (Jenkins

[39] Note that "status" is used in a very generic sense here. It is also possible to categorize different varieties according to the constitutional status of English in individual countries, while discrepancies between legal status and actual usage of English may exist, of course, and some countries that use English as *de facto* official language (e.g. Great Britain and the USA) do not have a constitutionalized language at all (see McArthur 1998: 38–42 for an overview of the respective official and actual status of English in a number of countries).

[40] Hence the labels "first diaspora" for these countries and "second diaspora" for the former colonies that are seen as members of the "Outer Circle" (Jenkins 2009: 18).

2009: 18) in the sense that they represent established and codified standards of the English language.

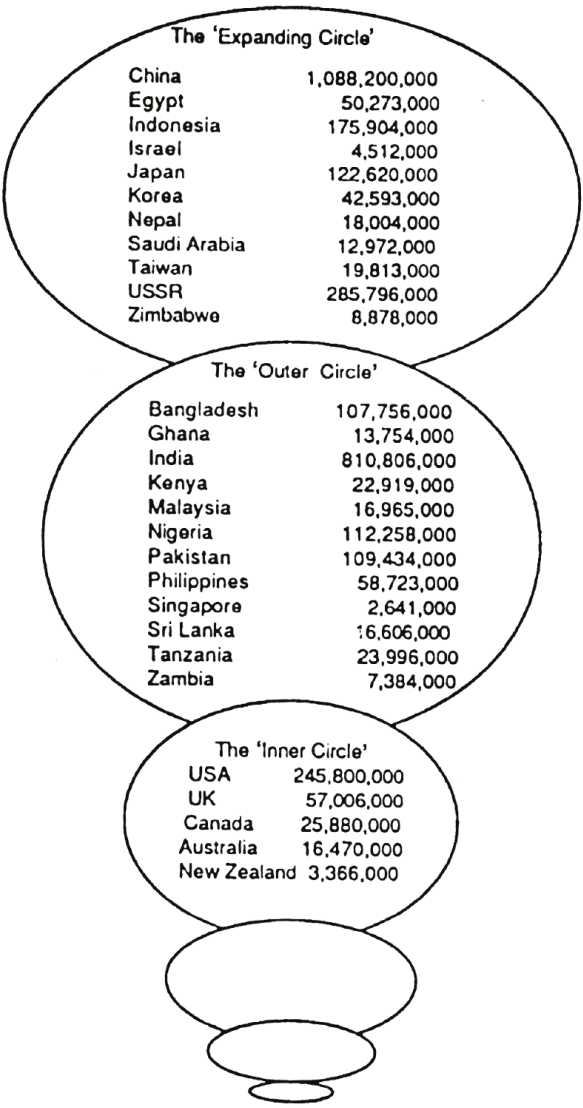


Figure 2.5.2. Kachru's (1985/1988) model of World Englishes

"Outer Circle" varieties are also found in former colonies, with the marked difference that there, English is assigned a "second language" (L2/ESL) or "non-native" status. Later characterizations argue that in these varieties, spoken norms may exist but that orientation toward the Inner Circle is still largely evident in written English (Mesthrie & Bhatt 2008: 29). Others take a more generic view, stating that these varieties are "norm-developing" and may have institutionalized English in some social contexts, for example as the language of the government or higher education (Jenkins 2009: 19–20). In contrast, the varieties of the "Expanding Circle" are typically "norm-dependent" (Jenkins 2009: 19) with respect to the Inner Circle varieties, and are commonly viewed as foreign language (EFL) varieties with no official status in the respective countries.[41]

Although the usefulness of this model and in particular its categorization into different variety types have repeatedly been acknowledged, some scholars see it as no longer adequate for the current situation of English. First, some speakers from the Outer Circle may in fact have English as their L1, and speakers from the Expanding Circle may move toward the Outer Circle as certain domains of their life (e.g. higher education or the workplace) become fully reliant on English. In some countries, this process is at such an advanced stage that transition from the Expanding to the Outer Circle seems imminent or may even have already taken place.[42] Second, Outer Circle speakers may extend their use of English beyond the officially recognized domains and might thus constitute a "grey area" (Jenkins 2009: 20) between the Inner and the Outer Circle. Third, the model does not acknowledge linguistic diversity within varieties (of any circle) (Mesthrie & Bhatt 2008: 31) and, finally, the model serves to uphold the "myth" (Pennycook 2007: 21) that varie-

[41] The source of the tripartite ENL/ESL/EFL model has been traced back to the work of Randolph Quirk et al. and ultimately to Barbara Strang's categorization into A-, B- and C-speakers (Strang 1970/Quirk et al. 1972; cited in McArthur 1998: 42–43; see further Schneider 2013: 134–135).

[42] Cf. Wächtler's observation that boundaries between ESL and EFL are fuzzy and more often than not based on the status of institutionalization in a country rather than on linguistic criteria (1977: 54–55). One case in point would be emerging European varieties of English (see e.g. Edwards 2010 on Dutch English).

ties are delimited by national boundaries.[43] As can be deduced, these criticisms are mainly due to static nature of the model and the fact that any model unavoidably abstracts from reality and is therefore always in danger of overgeneralization. Thus, Leitner suggests that the different variety types should not be seen as categorical but rather as "points (or areas) on a multidimensional scale" (1992: 186).

Revised versions of Kachru's model have since been developed (e.g. Graddol 2006: 110)[44] and the applicability of their categories demonstrated, for example by Kortmann & Szmrecsanyi (2011). In order to do more justice to the different ways in which the Inner Circle varieties came into existence from a sociohistorical point of view, they distinguish between "transplanted L1 Englishes or colonial Standard varieties" (such as NZE or AusE), "shift varieties" (such as IrE) and "Standard varieties" (such as BrE), and contrast these "high-contact" L1 varieties with non-native L2 varieties (as well as traditional "low-contact" L1 dialects) with respect to overall analyticity and syntheticity. The results show clear distinctions between the different types of varieties (Kortmann & Szmrecsanyi 2011: 281–283). This subcategorization of L1 varieties will be tested in the present study.

[43] For further criticism and alternative models, see McArthur (1998: 43–46), Bruthiaux (2003) and Jenkins (2009: 20–21). Jenkins adds that the model does not consider the influence of the individual speaker. She refers in particular to fully bilingual (and multilingual) speakers and speakers who are "non-native" by definition, i.e. due to their membership to one of the Outer or Expanding Circles, yet who are actually more proficient than some "native" speakers. See also Conrad, who problematizes the terminology "second language" (1996: 22), and Görlach, who criticizes some of the weaknesses of Kachru's model but defends the usefulness of the distinction between "native language", "second language" and "foreign language" (i) as necessary to allow an abstract conceptualization of the inherently complex issue of assigning a status to individual varieties of English (2002: 9) and (ii) for lack of a suitable alternative, given that "[t]here is as yet no objective method of determining a person's status as a speaker of English" (2002: 5).

[44] Graddol maintains that "Kachru himself, [sic] has recently proposed that the 'inner circle' is now better conceived of as the group of highly proficient speakers of English – those who have 'functional nativeness' regardless of how they learned or use the language" (2006: 110), a view that has subsequently been disputed as not fully mirroring Kachru's ideas (Jenkins 2009: 23–24).

Two other circle models (see Figure 2.5.3) that are often mentioned together due to their similar design are those developed by McArthur (1987) and Görlach (1990).

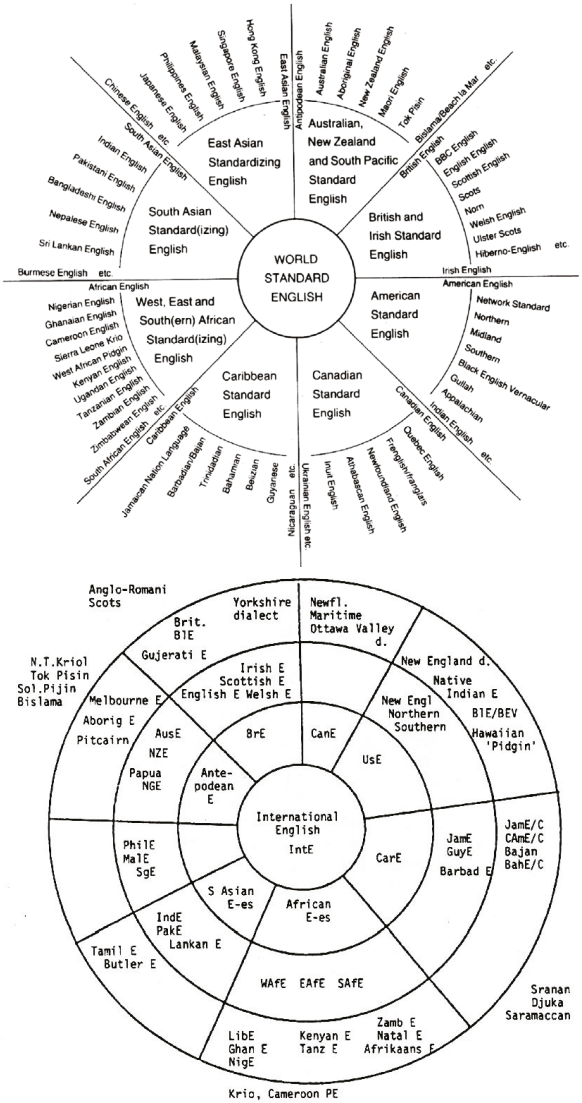


Figure 2.5.3. McArthur's (1987) and Görlach's (1990) models of World Englishes

Both have a non-regional variety (World Standard English/International English) at their center, which is to be seen as more of a theoretical construct than an actual variety of English.[45] The inner layer represents established and emerging regional and national varieties such as BrE, AmE, Antipodean English, African English, South and East Asian English, etc., some of them arguably again being idealized constructs. The outer layer in McArthur's model comprises various native and non-native varieties as well as pidgins, creoles and typical learner varieties (such as "Japanese English"), while Görlach distinguishes between "sub-regional semi-standards" and another layer of "non-standards" (such as traditional dialects), which are complemented by pidgins, creoles and related languages that lie outside the circle. In contrast to McArthur's and Kachru's models, Görlach does not include varieties with "foreign language" status. Note also that while McArthur's model neatly separates the varieties into eight geographical spaces, Görlach's version leaves some gaps, for example with regard to an Inner Circle category for East Asian Englishes, while neither model recognizes European varieties of English (such as Maltese English; cf. Krug et al. 2012). Like Kachru's representation, these circle models are static as well and rely mainly on geographical relationships and institutionalization status to establish their layers and categories.

[45] A related predecessor is the functional (not circle-shaped) model developed by Wächtler (1977: 51). He categorizes varieties of English according to their area of circulation and differentiates between international varieties (such as international scientific English), territorial varieties (such as BrE, AmE, IrE, AusE), regional and subregional varieties (such as Northern BrE, Southern AmE) and "foreign contact" varieties (such as Jamaican Creole). He makes some apparently surprising categorizations, such as "English as a Second Language" as an international variety. While the proposal of the existence of a uniform ESL variety would seem to be an oversimplification, it is purposeful, anticipating the later finding of shared common features in many varieties (despite their physical distance and range of potentially influential substrates; cf. e.g. the contributions in Kortmann & Lunkenheimer 2011). Wächtler also includes a variety called "Standard West Indian English" as territorial variety. It has to be noted that, although such a variety does not exist in any codified form, the terminology implies that geographic proximity plays a decisive role in shaping varieties (see further Section 2.5.3).

Modiano (1999) takes a radically different approach,[46] positing a circle model in the widest possible sense, but with overlapping circles that constitute a "common core" of English (as an "International Language", hence the label "EIL" for this area). As can be deduced from Figure 2.5.4, his representation relies on shared linguistic features rather than on geographical factors.

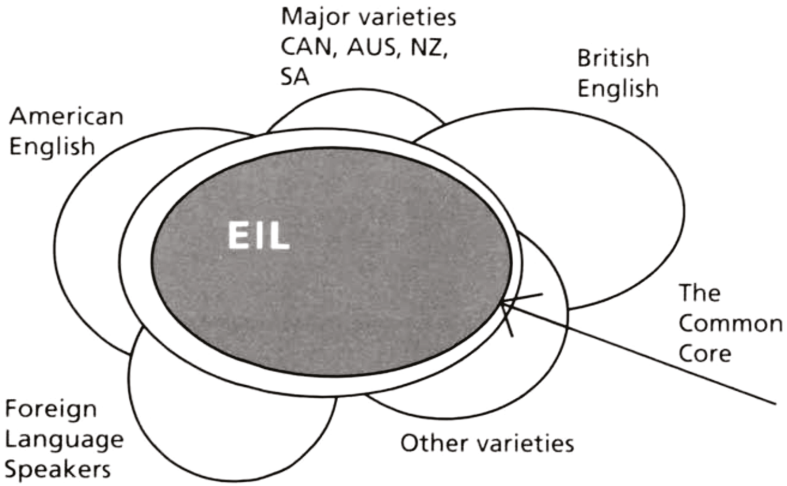


Figure 2.5.4. Modiano's (1999) model of World Englishes

In detail, the shaded central EIL area is constituted by what is accessible to both (competent) native and non-native speakers, while the white central area contains features that are in transition away from or toward international comprehensibility (Modiano 1999: 11). The surrounding (and partially overlapping) circles are categorized as five groups of varieties with idiosyncratic features[47] that are not intelligible to speakers of other varieties located in the areas incongruent with the core area. This model fundamentally differs from those presented above in that Modia-

[46] For a description of the genesis of this model see Jenkins (2009: 21–22). A similar approach with a structural "Common English" component was proposed earlier by Algeo (1988; cited in Leitner 1992: 180) and is also reflected in the "core" versus "periphery" approach mentioned by Quirk et al. (1985).

[47] For a list of examples of such features, see Modiano (1999: 8–9).

no conceives of the status of the individual varieties in a non-hierarchical manner and does not assign any type of variety as more or less central to the core area (although it may be argued that BrE is seen as the most influential as it covers the largest area). In addition, the model is synchronic and does not take sociohistorical developments into account, although dynamic changes could be incorporated by way of changing circle sizes and overlap.

Unlike Modiano's (1999) model, Schneider's (2007) approach toward transplanted Englishes in the widest sense ("postcolonial Englishes" in his terminology) is mainly historical, but also predictive in that he develops a uniform phase model for varieties of English. Schneider works with the underlying assumption that essentially similar processes (e.g. language contact processes; cf. 2007: 89)[48] are responsible for the form and status of English in various societies, and that all varieties can be located in one of the five abstracted phases (see Figure 2.5.5) based on various criteria.

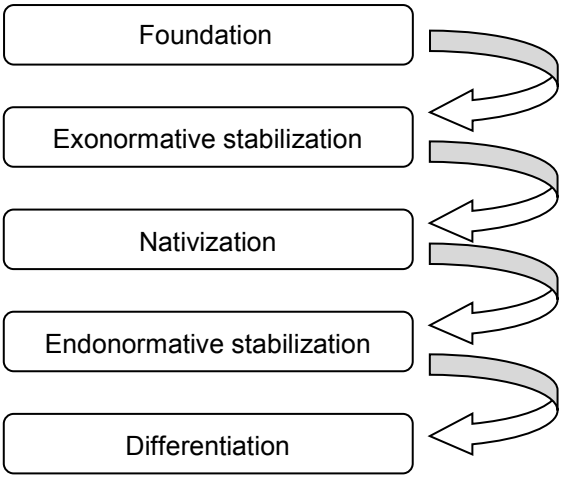


Figure 2.5.5. Schneider's (2007) dynamic model of World Englishes

Schneider offers an alternative to the conventional models that categorize varieties according to the role they play in different countries by also

[48] Thomason (2001) provides an in-depth introduction to language-contact phenomena.

including the perspective of the indigenous population and by considering issues of identity and identity construction among the affected social groups. This perspective also enables him to consider the implications of multilingualism, which is not possible in other models (Jenkins 2009: 20). A more precise description of the five phases is as follows (Schneider 2007: 33–55):

(i) Foundation (phase one) sees the introduction of English into a society in a particular geographical area, often for reasons of settlement or trade. Language contact occurs both between English and the local languages, and between the various dialects of the settlers. The first type of contact is likely to lead to the emergence of a pidgin, while the second process eventually leads to the establishment of an immigrant koiné. At this stage, only a few speakers are bilingual and, from a structural perspective, mutual borrowings are usually restricted to vocabulary items (e.g. place names, terms for particular botanical and faunal phenotypes).

(ii) Exonormative stabilization (phase two) refers to the stabilization of both the political situation (British domination)[49] and the language (orientation toward BrE rules and norms). In colloquial usage, however, the immigrant koiné becomes more and more established. An indigenous elite with knowledge of English emerges, resulting in an increasing number of bilingual speakers (through increased contact between the settler and indigenous communities or through education). Structurally, local lexical items are increasingly adopted by the settler communities, while the first instances of "structural nativization, [that is,] the emergence of structures which are distinctive to the newly evolving variety" (Schneider 2007: 39–40) are observed, which can be viewed as the first steps of a creolization process.

(iii) Nativization (phase three) is a decisive transitional period in which both political and linguistic allegiances, potentially due to a defining historical event, shift away from the traditional orientation toward Britain and a new, local identity is recognized. This usually

[49] The same naturally applies to varieties stemming from AmE.

goes hand in hand with a diminishing importance of political, cultural and socioeconomic differences between the original settler and indigenous groups, as well as with extensive efforts at English language acquisition on the part of the indigenous population. By the same token, local pronunciation, lexis and patterns of grammar usage, which may be influenced by learner effects, substrate features and characteristics of the koiné, start to stabilize in speakers of both the indigenous and (less so) the original settler communities. In this formative period, the stabilizing local usage is often contrasted to the (once) more prestigious forms of BrE and complaints about the "deteriorating" state of English in the respective community are heard. From a structural perspective, "negotiation" (Thomason 2001: 75) results in a convergence between the groups of the former L1 and L2 varieties. Further characteristics of this phase are code-switching and the emergence of a "mixed code" (Schneider 2007: 47).

(iv) Endonormative stabilization (phase four) is typically accompanied or preceded by the establishment of a strong local identity or nationhood (involving both the original settler and indigenous communities) that reflects independent and local sociocultural and political realities, and is thus distinctly non-British. A further by-product is the emergence of literature in the local variety. The acceptance of this new identity is reflected linguistically in a positive stance toward local forms and norms, illustrated by initiatives for the codification of local lexis and occasionally also of grammar. This supports claims that the new variety is a self-sufficient and homogeneous entity (Mesthrie & Bhatt 2008: 35). As a rule, the vast majority of erstwhile indigenous speakers have shifted to English in this phase, some of them retaining distinctive characteristics in their speech.

(v) Differentiation (phase five) is the stage at which the society no longer defines itself in contrast to the former colonial power, but rather sees the social and linguistic community as an independent whole made up of various (social, regional or ethnic) components. As the name of this phase implies, the former koiné differentiates over time into a range of social and regional dialects.

Schneider further suggests that colonization type may be a factor in shaping the individual varieties under investigation.[50] As a general rule, more non-standard patterns are found in "settlement colonies", such as American, Caribbean and Australian regions, than in "exploitation colonies", such as African and Asian regions (Schneider 2007: 84–85; see also Schneider 2013: 136–140).[51] If colonization type effects can be observed not only in non-standard patterns but also in the area of standard usage (e.g. of the PrPf), this would imply that varieties of the same colonization type should use standard constructions in a similar fashion and group together, as should varieties that are located within the same, or at least an adjacent, phase of the model.[52]

---

[50] Other features and processes that may influence the development of postcolonial varieties of English not discussed here for reasons of space are language policy, adstrate languages, creolization and the association of English with elitism (see Schneider 2007: 57–68 and the introductory sections for each variety in Chapter 5 for detailed analyses).

[51] Due to the manifold underlying processes and implications with regard to the different types of colonization, other authors have doubted the validity of a uniform path of development for both transplanted L1 and L2 varieties (e.g. Mesthrie & Bhatt 2008: 35; see also Evans 2009: 279). In addition, Schneider's model has been criticized for overemphasizing the importance of political events and neglecting language-internal developments (Mesthrie & Bhatt 2008: 35–36). See further Hickey (2004: 13), who argues that the "founder principle", i.e. the influence of different dialects that provide the initial input for a new variety, is of paramount importance for the later shape of a new variety of English. Following that view, Schneider underestimates the impact of the different types of koiné that emerge. In a similar vein, others have pointed out that the dynamic phase model does not sufficiently explicate the characteristics of the phases one and two (Evans 2009: 280).

[52] Cf. Brutt-Griffler, who argues that the individual situation of every colony ("local conditions") may have led to a spread of English with different outcomes (2002: 87). In addition, she draws attention to the fact that even if L2 varieties possess a full functional range, the setting in which the language was acquired is markedly different from L1 countries (2002: 134). This suggests (i) that the different types of acquisition are reflected in the usage of particular constructions, and (ii) that clear dividing lines should emerge between varieties of different types (traditional L1 vs. transplanted L1 vs. L2). Note, however, that from a terminological point of view, Brutt-Griffler prefers a distinction between mono- and bilingual speech communities rather than "native" versus "non-native" (2002: 137). Further, there should be marked differences between former British and American colonies, given that (i) unlike the US in the Philippines, the British empire had no uniform colonial administration and language planning policy before the 1920s, and (ii) "American imperial policy insisted on teaching English to every student, [while] the British policy limited the number of the students exposed to the formal teaching of English to meet the local de-

An unconventional model, which as yet has not found its way into the wider discourse of World Englishes studies, was developed by Minnich (2004). She rejects the sole orientation along geographical, historical and political lines that is characteristic of most models, where "political space is taken to coincide with social space" (Minnich 2004: 199). She argues that these linguistic approaches, unlike models from many other academic fields, such as economics or sociology, fail to take into account the implications of globalization (see Section 2.5.2), such as the "declining empirical validity [of the] putative congruence of language, nation and culture" (Minnich 2004: 200). Therefore, inspired by socioeconomic approaches and in particular by Appadurai's (1996) "social imaginary" model and its five dimensions of cultural flow, she proposes an alternative breakdown into "post-territorial scapes" (Minnich 2004: 208). Specifically, she recognizes "ethnoscapes", that is, groups of people who are highly mobile and share a form of mediated culture rather than actual physical proximity; "technoscapes" in the area of informational and mechanical technology; "financescapes" of global financial activities; "mediascapes", that is, implications of the different ways of producing media content electronically; and "ideoscapes" of ideas and images in the public sphere (Minnich 2004: 207–209). While Minnich herself concedes that the boundaries of and between these scapes are not easy to define and it must be stated that some of the ideas are developed only to a rudimentary extent, particularly the first concept, "ethnoscapes", may be relevant for the study of World Englishes (see also Mair 2013). It takes account of various global cultural and media influences, and can therefore be seen as an overriding factor when physical or (putative) cultural proximity between two varieties or the phase history of one variety do not explain a particular structural development. In an extreme case, this would suggest that physical proximity and historical relations between

mands for English-educated subjects of the empire" (2002: 89). This suggests that PhiE patterns of usage should diverge from those of other varieties in the same region. Fishman also claims that there may be substantial differences between former British and American colonies: "[i]t is certainly high time that someone examined whether countries in the former American colonial orbit and countries in the former British colonial orbit have turned out the same way or not *vis-à-vis* their respective metropolitan varieties, since British imperialism and American imperialism differed so importantly" (Fishman 1996a: 9).

varieties play no part in a globalized world at all. A more conservative interpretation would suggest that similarities between physically and historically distant varieties can at least partly be explained in this way (in combination with general processes of language acquisition and contact).[53]

As is also clear from the foregoing, many varieties emerged in environments where issues of language contact played and still play an influential role. In this regard, some further considerations from this field are therefore in order, although it is acknowledged that the majority of the models presented in this section take account of language contact in one way or another. Theoretical works on contact-induced language change mention both external/social factors (such as intensity of contact between the speakers of the two languages, linguistic attitudes and imperfect learning) as well as internal/linguistic factors (such as typological distance between the languages involved or universal markedness) as predictors for the degree and kinds of change that may occur. At the same time, it is evident that a combination of both internal and external factors is typically responsible for a particular outcome (Thomason 2001: 60–62; see also the list in Hickey 2004: 1–2).

Two basic "shift" scenarios toward a target language are established, depending on the intensity of the linguistic integration in the community (explained below for contact situations with English as target language; cf. Schneider's colonization types introduced above). The first scenario is that shifting speakers are not integrated and therefore constitute a separate ethnic or national group. This usually implies that features of the native language(s) of the shifting speakers are transferred to and remain fixed in their version of the target language (observable e.g. in IndE). The second scenario is that the shifting speakers are fully integrated into the community of the speakers of the target language, which results in an altered version of the target language containing some elements of the shifting variety (observable e.g. in IrE; Thomason 2001: 75). The second scenario is particularly likely if the shifting group is

[53] It is evident that a monocausal explanation as a rule does not adequately mirror linguistic reality; therefore, the usefulness of the concept of scapes but also its limitations should be recognized.

relatively large compared to the number of target language speakers (Thomason 2001: 78–79). This holds for all L2 varieties of English and predicts that they should be structurally different from L1 varieties of all types. Other relevant observations in this respect are (i) that the outcome of contact between typologically distant languages should be different compared to typologically close languages (Thomason 2001: 77), and (ii) that languages that are geographically close often share common features (Thomason 2001: 104; see also Muysken 2010: 265). Again, for English, (i) would imply that all L2 varieties should have markedly dissimilar patterns compared to both traditional L1 and transplanted L1 varieties, where the substrate was either typologically close (e.g. Irish and English) or where contact with indigenous languages only played a marginal role (e.g. AusE, NZE, CanE), while (ii) suggests that geographical proximity (e.g. in the case of Jamaica with respect to the US and Canada) may override traditional political or historical relationships that are reflected in language.

In sum, although the hypotheses and findings from this area are only partly applicable to World Englishes, language-contact studies complement other general models of World Englishes by offering patterns of explanation for a wide variety of usages. The strengths and weaknesses of all of these approaches, however, still need to be tested with the help of empirical data.

2.5.2 Globalization of culture and language

It became clear in the preceding section that there is an area of conflict between static and dynamic models of World Englishes (which are mainly based on geographic, historical, political or sociolinguistic – that is, mainly identity-related – considerations) and models that view these approaches as no longer adequate as they fail to take implications of globalization into account. By the same token, it has been asserted in studies of World Englishes that comparing other varieties to BrE alone is no longer adequate in light of linguistic reality and that a more global (in the truest sense of the word) approach is needed (Hundt et al. 2004: 560). This approach should take account of the "fourth crossing" (Mes-

thrie & Bhatt 2008: 24) of English; that is, the spread of English through globalized culture, media and education.[54] This section provides a working definition of globalization and investigates the potential repercussions that a globalized culture may have on language and in turn on linguistic analyses, particularly of grammatical structures.

While older definitions of (linguistic) globalization are restricted to the possibilities offered by technology in transcending borders (Wächtler 1977: 61), according to Bhatt, globalization as a combination of macroprocesses "represents a new, post-traditional order, forging new identities, institutions and ways of life. It is 'the way we live now', in a worldwide network of social relations, *seemingly* unfettered by the constraints of geography" (Bhatt 2010: 94; emphasis original). Similar views, although with slightly different emphases, are taken in other recent definitions by Elliot, who states that "[g]lobalization is about border crossings – whether on foot or by keyboard" (2010: xvii), and Vaish, who identifies a "high level of connectivity in this phenomenon between nations, corporations and individuals" (2008: 30) as a key property.[55]

Despite this increased connectivity and permeability as well as despite intensified global flows of goods, people and culture, others take a more cautious stance, arguing that geography and issues of nationhood and identity are still central factors. As Bhatt (2010: 103) writes, "globalization is the complex and multifaceted interaction of localism and globalism" (see also Pennycook 2007: 26–30). For linguistic studies, this complicates the picture considerably, but at the same time offers the opportunity to test whether some of the older boundaries and categories are still valid or whether globalization indeed functions as a leveling device, as stated by Ramanathan et al. (2010: xvii). If the latter hypothesis turns out to be correct, this would imply that language is indeed deter-

[54] This carries some implications that give English a special status in contrast to other languages. Suffice it to say here that some of the principles that usually apply, such as the determination of structural features by the area in which a language is spoken (Muysken 2010: 265), can be discounted for English (but presumably not for its varieties).

[55] Particular importance is assigned to the domain of the economy, which may be relevant since English is also commonly seen as the prime language of the globalized economy (Brutt-Griffler 2002: 48–50). Note that, contrary to some claims, it has been stated that English as a global language rarely acts as a "killer language" but rather fills a functional gap in a multilingual environment (Sharma 2001: 345).

mined by, or at least related to, economic, political and cultural macro-processes (Bhatt 2010: 94), which would potentially result in the need to re-assess the established models and concepts.

In contrast to the relatively neutral stance toward globalization that is inherent in the definitions given so far, other definitions see globalization as a more or less unidirectional (and overall negative) development. For example, Elliott (2010: xxii) regards it as "the Westernization, or Americanization, of the world."[56] This terminology suggests a strong influence of (Anglo-)American culture on the cultures of other communities and nations (Phillipson 2009: 152; see also Hartford 1993: 1), which has been applied in the area of language explicitly with the coinage "linguistic imperialism" (Phillipson 1992; see also Leitner 1992). While some have criticized the concept for merely transferring political and economic power relationships to the domain of language (e.g. Conrad 1996: 25), it nevertheless establishes an interesting test case for linguistics insofar that it draws attention to the fundamental "tension between cultural homogenization and cultural heterogenization" (Phillipson 2009: 30) that is inherent in globalization. For linguistic studies, acceptance of the homogenization hypothesis would again imply current or at least future convergence of varieties of all kinds, but with an orientation toward AmE (and less toward BrE).[57] The latter position has been dismissed as unlikely by other authors, however, who maintain that, although forms are taken over from AmE, their functional usage in particular varieties may still differ substantially from American patterns (Leech et al. 2009: 258).

---

[56] Other, more metaphorical terms referring to globally successful American brands, which are used interchangeably, include "Coca-colonization" or "Walmarting" and, with more specialized meanings, "Disneyfication" or "McDonaldization". All of these carry a pejorative connotation.

[57] Allerton presumes that the orientation toward AmE rather than BrE is due not just to the political and economic power of the US, but also to the notion that AmE is more "affected by the number of non-native speakers" (2008: 42). However, this line of argumentation is not supported by evidence (such as numbers of non-native speakers of English in the US). Other researchers have argued that the actual influence of BrE is bigger than could be suspected from Britain's political and economic weight due to its success in marketing its "brand" of English worldwide, especially through the British Council (Dua 1996: 578–584: see also Schmied 1996: 184).

Yet another terminological concept, "anglification" (Fishman 1996a; 1996b), is also relevant with regard to globalization and English. Fishman establishes seven criteria – or rather, domains of usage – where English is introduced (namely, elementary education, tertiary education, print media, non-print media, the technology/commerce/industry complex, governmental services and operations, indigenous informal usage), and argues that the anglification of a country can be measured according to the number of anglified domains (Fishman 1996b: 634).[58] Therefore, Fishman's approach could be classified as being primarily geographical. But he also takes issues of social stratification into consideration and concludes that anglification is most prominent among higher social strata and in domains that as a rule are dominated by members of the elite, such as tertiary education, the technology/commerce/industry complex and, to a lesser extent, print media and governmental services and operations (Fishman 1996b: 636–637). This implies (i) that patterns of usage by speakers of a comparable social background should be relatively homogenous, (ii) that these speakers should use mostly standard patterns, and (iii) that particular genre types, such as print media texts are relatively homogeneous and exclusively use standard patterns.

As a final point, I now examine the implications of globalization for the study of English grammar. On the one hand, an internationalization of grammatical patterns has been identified; that is, a "convergence of regional varieties with the predominance of AmE variants" (Hundt et al. 2004: 588),[59] which tallies nicely with some of the more general claims outlined above about the nature of globalization. On the other hand, there is a more general feeling of discontent among some authors that both theoretical and empirical descriptions of varieties of English still focus too heavily on contrasts with AmE and especially BrE. Because

[58] Fishman's model also provides the opposite perspective, establishing a "total anglification resistance score"; that is, the more restricted the English usage is in more individual domains, the higher the score.

[59] See also Wächtler (1977: 67), who observes only a low amount of divergence as regards pronunciation in L2 varieties, given that British, American and Australian models are viewed as influential in this domain for particular geographical regions. It remains to be seen whether these findings are transferrable to the area of grammar.

the differences (i) between the "old" L1 varieties and other varieties[60] and (ii) across all varieties in the areas of syntax and morphology are usually gradual rather than categorical (Hundt et al. 2004: 560), such "deficit linguistics" analyses are seen as inadequate.[61] I would therefore like to suggest that varieties of English be evaluated in a manner that is as inclusive as possible, incorporating varieties of all types for which comparable material is available.

With regard to the usefulness or applicability of the models of World Englishes listed in the preceding section, a recent empirical study (Hundt & Vogel 2011) comparing the distribution of one structural feature (the progressive) found that patterns of usage across a number of varieties are relatively unpredictable and apparently random in the sense that, when it comes to individual structural aspects, a neat categorization into L1, L2 and learner varieties is too simplistic (cf. also the related points in Omoniyi & Saxena 2010: 5 and the "Variety Spectrum" approach sketched in Bongartz & Buschfeld 2011: 45–48). Hundt & Vogel concluded that (i) globalization may be a potential cause for this blurring of boundaries and (ii) these categorizations might be more valid from a sociolinguistic than a structural point of view. In addition, their data show that "exotic" patterns are comparatively rare and support the hypothesis that varieties of English are largely convergent as far as individual grammatical structures are concerned. Whether these findings can be extended to other grammatical variables, such as the PrPf and its usage patterns, has not yet been tested.

2.5.3 World Englishes and versals

A relatively recent approach that originated in the field of linguistic typology is the search for *versals* (see e.g. Chambers 2003); that is, charac-

[60] Wächtler (1977: 69–71) may be seen as one of the first to move away from giving BrE (and, with certain caveats, AmE) a different status in contrast to other (transplanted) L1 varieties. Due to historical processes and developments (colonialism, globalization, etc.), he asserts that Great Britain has lost its former "ownership" of and "monopoly" on the English language (1977: 111–112).

[61] Hundt et al.'s statements refer to the situation for NZE, but can be seen as valid for all varieties of English (see e.g. the contributions in Hickey 2012).

teristics or structural features that are shared across languages (or varieties). This is of interest in the present study as it provides a descriptive framework that may explain similarities or seemingly unmotivated connections between Englishes of different variety types (see Section 2.5.1) or that are geographically distant. Therefore, versals theory can be seen as an auxiliary approach that complements the more general views on globalization as discussed in the preceding section, in that it provides theoretical concepts whose validity can be tested with the help of corpus data.

Szmrecsanyi & Kortmann (2009) provide an overview of typological versals. For the present study, of particular importance are the concepts (i) *areoversals*, that is, "features common to languages which are in geographical proximity to each other" (Szmrecsanyi & Kortmann 2009: 33), and (ii) *varioversals*, that is, "features recurrent in language varieties with a similar socio-history, historical depth, and mode of acquisition" (Szmrecsanyi & Kortmann 2009: 33). I would like to suggest that with regard to (i), "languages" could be substituted by "varieties" in order to test the special case of English and its pluricentric nature (Hundt & Biewer 2007: 249). In addition, also the concept of *angloversals*, that is, "features that tend to recur in vernacular varieties of a specific language" (Szmrecsanyi & Kortmann 2009: 33; English in the present study) could be put to the test by dividing the data into written and spoken sections, where the spoken sections would yield putatively more evidence. The latter point is worth mentioning in light of the fact that varieties may share features that cannot be explained by way of either (i) or (ii) (see e.g. Winford 2009: 225; Davydova et al. 2011: 292).

Again, however, it has to be acknowledged that reality may be much more complicated than suggested by a neat theory, and the generalizability of one feature or another across variety types and geographical regions has to be tested on a case-by-case basis (Szmrecsanyi & Kortmann 2009: 49). A follow-up empirical study of non-written data by the same authors revealed that variety type (varioversal) rather than geographical proximity (areoversal) is "the better predictor of overall similarity or distance between individual varieties" (Kortmann & Szmrecsanyi 2011: 276; see also Szmrecsanyi 2012), although the situation might be

converse or at least more complicated for individual features (see e.g. Hundt 2009; Hundt & Vogel 2011).

Others have advised even more caution as regards the usefulness of versals and have suggested that it is more adequate to think in terms of "degrees of similarity" (Sharma 2009: 173) of varieties of World Englishes, first and foremost excluding substrate explanations before invoking a versal explanation, as similar patterns in two varieties may in fact be due to two types of substrate influence that are typologically unrelated (Sharma 2009: 192; see also Winford 2009: 225 and Hickey 2010b: 21).[62] The converse argument, then, is that if substrate influence can be eliminated (and many researchers doubt that this is feasible at all), a versal feature may be identified. Therefore, as many varieties as possible should be compared to minimize the likelihood of many substrates exerting a comparable influence that could be misinterpreted as a versal feature.

[62] Another criticism of typological versals as established by Szmrecsanyi & Kortmann (2009) is that their approach is overly abstract in the sense that it only recognizes the presence or absence of a feature when determining the similarity of two varieties without considering "semantic or other conditioning" (Sharma 2009: 171). This issue is addressed in the present study (although only for one particular grammatical area and not to establish an overall measure of similarity between varieties as intended by Szmrecsanyi & Kortmann 2009).

3 Previous research

Having presented a fairly general framework for the study of the English PrPf, this chapter now provides a survey of relevant previous research on the topic. The three principal motivations are (i) to update and complement literature reviews from older studies, as the PrPf represents a dynamic area of research with a never-ending flood of new publications, (ii) to systematize previous accounts, and (iii) to provide points of reference for the present study.

Specifically, in the following sections the PrPf as introduced in influential reference grammars is considered, and different standpoints regarding the categorization of the PrPf, which revolve mostly around the tense-aspect debate referred to above (Chapter 2), are addressed. Furthermore, a synopsis of relevant theories and studies on the seemingly inexhaustible subject of the semantic and pragmatic interpretation of the PrPf is provided. Findings with respect to the PrPf in non-standard varieties are also compared, and the results from previous corpus-based research discussed. Finally, the implications for the empirical part of the present study are outlined.

3.1 The Present Perfect in descriptive grammars[63]

Two reference grammars that have been very influential and are currently widely used are the *Comprehensive Grammar of the English Language* (Quirk et al. 1985) and the more recent, corpus-based *Longman Grammar of Spoken and Written English* (Biber et al. 1999).[64] Therefore, it is worth considering the statements on the PrPf contained in these major works. Both recognize two actual tenses (present and past) and accordingly categorize the PrPf under aspect (Quirk et al. 1985: 4.18; Biber et al. 1999: 452). Quirk et al. concede that "aspect is so closely connected in meaning with tense, that the distinction in English grammar between tense and aspect is little more than a terminological convenience which

[63] An extended version of this section can be found in Werner (2013a).

[64] For a discussion of the PrPf in older grammars, see Vermant (1983: 12–15).

helps us to separate in our minds two different kinds of realization: the morphological realization of tense and the syntactic realization of aspect" (1985: 4.17). Thus, in their view the defining criterion for excluding the PrPf from the tense category is a formal one in that only tenses can be structurally (i.e. morphologically) marked, an approach shared by Biber et al. (1999).

With regard to the distribution of forms with and without perfect aspect, Quirk et al. state that a perfect notion can be found in around 10% of all VPs (1985: 4.18), a finding by and large reproduced in the data analysis by Biber et al., although differences between individual registers exist. Fiction and news emerge as relatively perfect-friendly in general, while the relative proportions of the PrPf in contrast to the Past Perfect are high in news and conversation. In addition, both claim that perfects are more frequent in BrE compared to AmE (again with differences between different registers; Biber et al. 1999: 461–463; Quirk et al. 1985: 1.24, 4.20).

Functionally, the two accounts take a slightly different approach. While both agree that the perfect aspect implies anteriority (Quirk et al. 1985: 4.18; Biber et al. 1999: 460), Quirk et al. (1985: 4.20) describe the general meaning of the PrPf as current relevance. Biber et al. (1999: 460), in contrast, emphasize the continuative and anterior properties of the construction. The latter refrain from further explanations of potential semantic readings, while the former can be seen as proponents of a polysemous account that lists three basic underlying meanings of the PrPf, namely (i) state leading up to the present, (ii) indefinite events in a period leading up to the present,[65] and (iii) habit or recurrent event in a period leading up to the present (Quirk et al. 1985: 4.20).

In addition, both descriptions repeatedly stress the fact that the PrPf is interchangeable with the SPst in some contexts, which strongly contributes to the "overlap of meaning between tense and aspect" (Quirk et al. 1985: 4.18; cf. 4.20) of the PrPf construction, while the difference in meaning between PrPf and SPst may be attributed to temporal adverbi-

[65] For this reading, three further implications are considered relevant: "(i) that the relevant time zone leads up to the present; (ii) that the event is recent and (iii) that the result of the action still obtains at the present time" (Quirk et al. 1985: 4.21).

als (Quirk et al. 1985: 4.20; Biber et al. 1999: 467).[66] Various factors that may determine a speaker's choice between the two forms are listed, such as a more direct relation of the action to the present through the PrPf (Quirk et al. 1985: 4.20), but the accounts remain vague at best, which is in all probability due to the nature of the works as general reference grammars.[67] Other weaknesses that have been criticized are the restricted variational focus (standard BrE and AmE only; Phillipson 2009: 31) and the purely form-based analysis (Nordlander 1997: 16–18).

The *Cambridge Grammar of the English Language* (Huddleston & Pullum 2002) takes a markedly different approach toward the PrPf. The authors also distinguish "two primary tense categories, preterite and present" (Huddleston & Pullum 2002: 125). However, in their wider definition of tense that includes compound tenses the PrPf is categorized as a "secondary past tense" (Huddleston & Pullum 2002: 116; see also Huddleston 1996), while they recognize that alternatives to their categorization, such as "perfect as aspect", do exist. With regard to terminology, they use "past tense" as an umbrella term for both SPst and perfect forms (Huddleston & Pullum 2002: 116). This seems rather odd, in particular as they state at a later point that the PrPf is indeed a combination of past and present and even that "in the present perfect [...] the primary tense is present" (Huddleston & Pullum 2002: 142), which is inconsistent with their general approach to grammatical categories (see Figure 3.1.2 below).[68] They further explain their "past tense" umbrella

[66] In contrast to the considerations presented by Biber et al. (1999: 467–468), temporal adverbials with the PrPf are comparatively rare (Schlüter 2002a). See Declerck (2006: 316–326) for an attempt at a more elaborate description of possible selection criteria in sentences without temporal adverbials. He advocates that "when, in a noninterrogative clause referring to a bygone situation, there is no time-specifying adverbial, both the preterite (= SPst; V.W.) and the indefinite present perfect are in principle possible", and that pragmatic "speaker focus on NOW or THEN" (2006: 322; emphasis original; cf. Pulgram 1987: 384) is the crucial underlying notion between using either the PrPf or the SPst.

[67] Note that Biber et al. (1999: 464–466) provide interpretations of their corpus data with regard to the most common verbs that occur in combination with the PrPf and the special status of *have got/gotten* and *have had* as highly frequent possessive constructions.

[68] Cf. also Huddleston & Pullum (2002: 143): "The possibility of having present time adjuncts like *now* or *at present* shows clearly that we have present time meaning as well as present tense form." In addition, it seems redundant to propose a binary distinction (primary and secondary tenses) only to abandon it on a lower level of analysis and to substitute

category by stating that anteriority is the common denominator in both forms (2002: 139). Their overall model stands in stark contrast to the views expressed by Quirk et al. (1985: 4.18) and Biber et al. (1999: 460). For Huddleston & Pullum, the defining feature of the PrPf compared to the SPst is that the former is non-deictic (2002: 140–143) while the latter is deictic and "doubly anterior" as it "locates the writing anterior to an intermediate time which is anterior to the time of speaking" (2002: 140). The PrPf is only anterior to the time of speaking (2002: 142).[69] The explanation of the criteria for choosing between the two constructions is again based on pragmatic notions, such as speaker focus on either past or present (2002: 143).

Functionally, Huddleston & Pullum provide a polysemous account. This recognizes a twofold view with continuative uses, which are claimed to be usually accompanied by temporal adverbials (2002: 141–142), and non-continuative uses, namely (i) experiential/existential, (ii) resultative, and (iii) recent past (2002: 143–146 for example sentences and elaborations on when these readings are possible).[70] Similar to Quirk et al. (1985), they identify a general notion of current relevance for the PrPf, although they provide no further details on this concept besides the fact that the PrPf is somehow concerned with the "time span up to now" (Huddleston & Pullum 2002: 143).

The figures underneath capture the main differences between Quirk et al. (1985) and Biber et al. (1999) on the one hand and Huddleston & Pullum (2002) on the other hand, both with regard to their general approach toward grammatical categories and the location of the PrPf.

---

it with an ad-hoc category (past tenses) that is not anchored in the general model. This once again highlights the rather elusive character of the PrPf.

[69] *Mutatis mutandis* (with a change of perspective from anteriority to posteriority), this view recalls Mugler's (1988: 235) approach, which states that the PrPf expresses simple posteriority while the SPst expresses complex posteriority.

[70] Note that Huddleston & Pullum acknowledge that these classifications are not necessarily mutually exclusive (2002: 143).

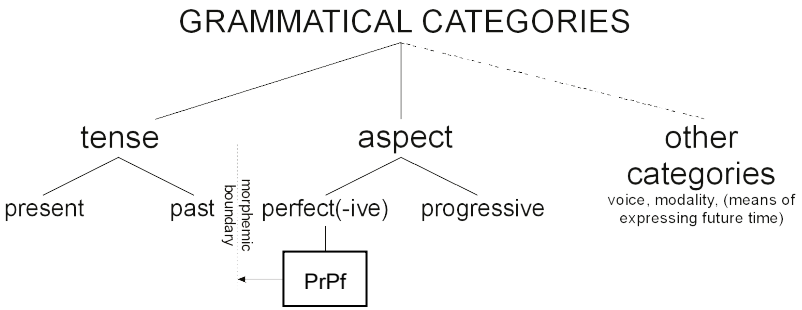


Figure 3.1.1. Grammatical categories as represented by Quirk et al. (1985)[71] and Biber et al. (1999)

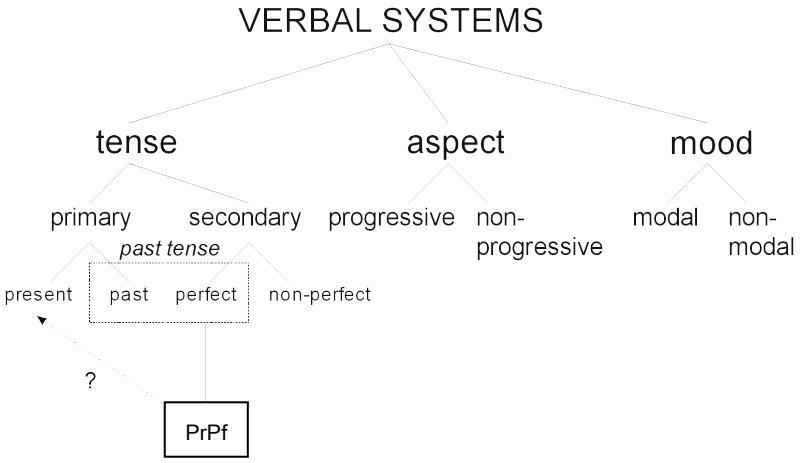


Figure 3.1.2. Grammatical categories (= verbal systems) as represented by Huddleston & Pullum (2002)

What is clear from these visualizations is that the PrPf does not fit into the general models without some kind of workaround. Therefore, while the two main approaches ("PrPf as aspect" vs. "PrPf as tense") represented in major reference grammars have both inherent merits and weaknesses, the important points for the current study are (i) that the

[71] Means of expressing the future are treated separately in Quirk et al. 1985 (4.41–4.48), while they are included under modals by Biber et al. (1999: 483–497), hence the notation in brackets in Figure 3.1.1.

PrPf indeed possesses features that are characteristic of both tense and aspect, which renders it an elusive category, (ii) that the PrPf may be viewed as creating reference to both present and past, and (iii) that there are different ways of categorizing the various functions of the PrPf. These inherent characteristics provide the fuel for the discussion below on the grammatical status of the PrPf in the theoretical literature.

3.2 Tense or aspect or something else?

As was shown in the previous section, a first hint that the English PrPf – or rather, the whole perfect paradigm – sits uneasily between the categories TENSE and ASPECT could be inferred from the fact that it is interchangeable with the SPst in some contexts (see also Koziol 1958: 502ff; Leech 2004: 35). This interchangeability has been covered extensively and expertly in other works (e.g. Elsness 1997 or Davydova 2011) and is only of secondary interest in the present study. In addition, it should be self-evident that TENSE and ASPECT are categories that closely interact in language in general and hence in English and its varieties (see the preceding section; cf. Platt et al. 1984: 67; Vasudeva 1985: 190; Salkie 1989: 31) to express *temporality*. Nevertheless, it is crucial to understand the reasoning behind the differing positions on the status of the PrPf as either ASPECT or TENSE (the two major positions), as a grammatical category of its own, or as a hybrid category[72] in order to be able to judge the adequacy of these theories in the context of World Englishes. The underlying hypothesis is that the vast majority of theoretical works on the PrPf base their findings on one (or both) of the "old" standards, i.e. BrE or AmE, and are therefore not fully adapted to the status or presumably even different statuses the PrPf could have in other varieties.

3.2.1 Perfect as aspect

The first major position to be outlined is "perfect as aspect" (as represented by e.g. Bauer 1970; Mugler 1988; Moens & Steedman 1988;

[72] For older reviews of theories on the categorization of the PrPf see e.g. McCoard (1978: 6–17); Fenn (1987: 245–249) or Panzner (1995: 17).

Salkie 1989; Bartsch 1995; Liszka 2003; Leech 2004; Michaelis 2006; Vater 2007 and Jaszczolt 2009). As noted above (see Section 3.1), this is also the approach favored by reference grammars such as Quirk et al. (1985) or Biber et al. (1999).[73]

Within the "perfect as aspect" paradigm, slightly different approaches can be observed; however, most accounts base their rationale on a negative definition, i.e. why perfects *cannot* be tenses in English. For example, in the minimalist, twofold view of the English tense system (see Section 2.2), it appears imperative that "perfect" be paralleled with "progressive", as both of these aspects can co-occur with verbs marked either for past or for present (Liszka 2003: 13). While this is doubtless true, the description remains at a solely form-based level and ignores functional facets.

Other approaches, notably the different versions of the temporal framework proposed by Klein (1994; 2009b), introduce an intermediate temporal level labeled *topic time*, "the time about which something is asserted (or asked)", which complements the *time of the situation*, the "time at which the situation obtains or occurs" (Klein 2009b: 46; cf. Klein 1994: 3–9). Here, aspect is seen as necessary to convey the connection between the situation and the topic time (see also Section 2.1), so it can metaphorically be described as "joint between clause-external and clause-internal structure" (Klein 2009b: 78). Two problems are posed by this analysis, which Klein himself identifies. The first one is that the exact difference between SPst and PrPf in English cannot be understood within the tense-aspect model alone (Klein 2009b: 53–54), so the introduction of a separate grammatical category "perfect" could be regarded as a remedy. The second problem follows from the potential solution to the first, and is of a formal-logical nature. Klein states that "if the perfect is also an aspect, it cannot be on a par with the perfective and the imper-

[73] All these accounts are implicitly based on the notion that the extensive formal tense systems of classical languages such as Greek and Latin are not transferable to modern languages such as English (Pulgram 1987: 382). Note, however, that the terminology (e.g. "perfect") has not altogether changed, although different grammatical concepts are described.

fective aspect, because these are found **within** the perfect" (Klein 2009b: 54; emphasis original).[74]

Yet others stress that there is a functional overlap between forms such as SPst and PrPf with regard to "basic temporal relations" (Michaelis 2006: 223), but that a distinction between tense and aspect is useful and valid as the former "merely locates reference time" while the latter "determines the manner in which the denoted situation relates to reference time" (Michaelis 2006: 241; for a similar view, see Bartsch 1995: 128).[75] If referring solely to past time was important, the PrPf as an alternative form to the SPst would probably not have emerged at all. This implies that a one-to-one congruence between the SPst and the PrPf, and hence the use of the PrPf as genuine tense (e.g. in the sense of the Quirkian grammar), should not be in evidence in linguistic data. A similar line of argument is used in theories that regard the characterization of internal developments of situations as a typical property of aspect that distinguishes it from tense, which connects different external situations to one another (Nordlander 1997: 162).

It has to be noted, however, that most of the accounts that see English perfects as an aspectual category repeatedly acknowledge tense properties (Nordlander 1997: 162; Leech 2004: 35; Petersen 2004: 116; Michaelis 2006: 223; Jaszczolt 2009: 86) and often give pragmatic reasons for the choice between the SPst and the PrPf (Michaelis 2006: 239).

3.2.2 Perfect as tense

Some researchers take the middle ground, stating that the PrPf cannot be categorized at all and should rather be placed "between tense and aspect" (Anderwald 2009: 49). This underscores the fact that the "perfect as aspect" view is by no means commonly accepted. As these middle

[74] *Imperfective* and *perfective* roughly correspond to the presence or absence of progressive marking in English (but cf. Section 2.2).

[75] See also Michaelis, who maintains that the PrPf "does not represent a unitary aspectual construction, but a complex of such constructions" (1994: 153), e.g. depending on the different semantic readings that are possible.

ground accounts often remain vague,[76] a description of different versions of the major alternative view, i.e. "perfect as tense", would seem to be a fruitful first step toward identifying the rationale behind such accounts (see Section 3.2.4 on alternative views). Note that, within this paradigm, the elusive nature of the PrPf is once again emphasized by the fact that this form has been categorized both as a present and a past tense. Both views are explained below.

A rather simplistic explanation of the PrPf as a present tense could be based on the finding that HAVE, the operator in PrPf constructions, morphologically represents a present tense form (Salkie 1989: 7).[77] This view echoes an earlier approach that sees the PrPf, with its two functions "retrospective present" and "inclusive present", as a link between past and present (Jespersen 1931: 47). Therefore, the PrPf should be functionally distinguished from the SPst (see also Kuhn 1989: 518). Others argue that the PrPf must be a present tense for reasons of analogy, as, like the Simple Present, it establishes reference to the moment of speaking (Ota 1963: 41). Unfortunately, those accounts that argue for the PrPf as present tense are unable to explain the special status of the construction, and have therefore not enjoyed widespread circulation or acceptance.[78]

---

[76] Cf. Wright's analysis of the PrPf as a construction that expresses "emotional anteriority" for a notable exception (1986: 75–76).

[77] See also Rastall, who claims that "[t]he writers of Standard reference grammars (both for native speakers and for English language learners) have all rightly stressed that in English [...] the present perfect is predominantly a present tense" (1999: 79). As was shown in Section 3.1 above, this claim is misleading, as reference grammars either categorize the PrPf as aspect (e.g. Quirk et al. 1985; Biber et al. 1999) or as a past tense (e.g. Huddleston & Pullum 2002). Furthermore, as will be shown in Section 3.3 below, the different uses or readings of the PrPf have also been hotly debated and are by no means universally agreed on, as maintained by Rastall (1999: 79–80).

[78] But cf. Reichenbach's label "Anterior Present" (1947: 297). Within the domain of tense logic, the discussion of his theoretical tense system (with the three constituents *event time*, *reference time* and *speech time*; Reichenbach 1947: 288) and its applicability to English and other natural languages has been persistent, as have suggestions for revised versions of that system (e.g. Bennett 1978; Comrie 1985; Fenn 1987; Salkie 1989; Binnick 1991; Klein 1992, 2009a, 2009b; Vlach 1993; Klein & Vater 1998; Michaelis 2006; Rothstein 2007; Vater 2007).

A diametrically opposed position within the "perfect as tense" paradigm sees the PrPf as a past tense. It is most notably advocated by Huddleston (1996) and has found its way into the *Cambridge Grammar of the English Language* (Huddleston & Pullum 2002) in a basically unaltered form (see Section 3.1 above). Hence, suffice it to repeat here that the rationale behind this approach is that anteriority to the moment of utterance[79] is the common denominator for both the PrPf and the SPst, and that the two forms are truth-conditional equivalents (Inoue 1979: 587).[80] However, they differ both formally, as the analytical PrPf is less grammaticalized than the inflectional SPst, and functionally, as the PrPf is non-deictic[81] and expresses complex anteriority (it can refer to situations wholly before or before and up to the moment of utterance) while the SPst is deictic and merely expresses simple anteriority (it can refer only to situations wholly before the moment of utterance).[82] This distinction is expressed in the labels "primary past tense" for the SPst and "secondary past tense"[83] for the PrPf (Huddleston 1996: 106–112; see also Ota 1963: 2).[84]

---

[79] Cf. also Weinrich's notion of *Rückschautempora* ('tenses for looking back'; Weinrich 2001: 88).

[80] Contrasting with this view is the approach presented by Wright, who states that anteriority, despite being the essential feature of the PrPf, does not differentiate the PrPf from the SPst as "[t]he anteriority marked by the perfect is not so much temporal as emotional in character" (1986: 85). This recalls pragmatic "speaker focus" considerations (Michaelis 2006: 239 and Section 3.1). Yet, and in spite of her discussion and recognition of some aspectual characteristics (Wright 1986: 72–76), Wright categorizes the PrPf as "closer to tense than to aspect" (1986: 70) due to its deictic meaning. Her notion of tense, however, seems to be unconventional or at least highly flexible. In the context of a contrastive analysis of the PrPf and the Past Perfect, she comes to the conclusion that the two forms "contrast in having different tenses" (1986: 74).

[81] Note that earlier accounts within the "perfect as tense" paradigm consider the PrPf as deictic, as it expresses a paradigmatic relation between the moment of the situation and the moment of speaking (Markus 1977: 51–59).

[82] Leaving aside the constraint that the PrPf does not usually combine with definite temporal adverbials, one weakness of this approach is that the functional scope of the PrPf covers the range of the SPst. In consequence, the SPst would be a redundant element of the grammatical system of English or at least restricted to certain registers (as can be observed in other European languages).

[83] A different but related concept is that of "function times" (Harder 1997: 500). Within this model, tenses such as "past" may have various function times on a subordinate level that do not provide genuine temporal information but rather different points of view. The

A third approach is to consider the PrPf as a separate tense that is not a subcategory or modification of either past or present. This approach is mainly based on the observations that unlike a typical aspect (such as the progressive) the PrPf does not solely represent the internal temporal constitution of a situation (Lindstedt 2000: 368), and the expression of current relevance, often considered the defining semantic feature of the PrPf, is not a sufficient criterion for its classification as aspect. First, current relevance remains a vague concept (Declerck 1991: 340) or may be only a meaning derived from other semantic labels (e.g. "anteriority"; cf. Wright 1986: 78) and, more importantly, depending on context is inherent in other forms such as the present tenses and arguably all tenses (McGilvray 1991: 48; Harder 1997: 382). Moreover, if the PrPf is categorized as aspect, it is not clear whether it represents past or present or both on the tense level. Thus – and as the main function of the PrPf is to express temporal anteriority, a typical property of a tense[85] – it must be a tense in its own right (Declerck 2006: 38, 109).

This reasoning has repercussions for Declerck's view of the English tense system as a whole, necessitating the introduction of an additional absolute time zone besides past and present. Declerck explains that what differentiates the PrPf from the SPst is that while the former locates a situation in the "pre-present zone"[86] the latter locates it in the

---

PrPf would thus be past (tense) + anteriority to the "time of reckoning" (function time) (Harder 1997: 382; see also Section 3.3.1.1).

[84] Lindstedt offers an alternative approach within the "perfect as past tense" model, claiming that the semantic functions of the PrPf play a major role with regard to its categorization. He argues that the PrPf in AmE, for example, is "more tense-like than its British counterpart" (Lindstedt 2000: 371, 378), given that current relevance, arguably the defining feature of the perfect as aspect, is not as consistent in AmE in comparison to other readings, such as experiential or indefinite past. Therefore, he suggests that the label "perfect" for the PrPf in AmE is inadequate (Lindstedt 2000: 378).

[85] Cf. also Dubois (1972). As has repeatedly been shown above and will be elaborated further below (see Section 3.3.1.1), other authors see the expression of anteriority as the central feature of perfects (and thus the PrPf) and not of past tenses in general (see e.g. Panitz 1998: 241; but cf. Portner 2003: 481–482 for a view that considers the pastness of the PrPf to be based solely on the Aktionsart of the verbs involved).

[86] An earlier version of Declerck's tense system does not operate with the "pre-present zone". In this version, the PrPf is situated in the "present zone" (Declerck 1991: 320). Kirsten (1994) and Panzner (1995) use the same "pre-present" label. Cf. also other terminology with similar implications, such as "prospective" or "anterior present". For a discus-

"past zone".[87] The existence of the "pre-present" as a time zone (and therefore the PrPf as separate tense) is grounded in the finding that the PrPf "expresses a temporal structure which is different from the semantic structure of any other tense" (Declerck 2006: 110; cf. Hantson 2005: 246)[88] and is therefore on a structural par with the other "absolute tenses" (Declerck 2006: 25),[89] *viz*. the present, the past, and the future.

3.2.3 Perfect as a grammatical category of its own

As shown in the preceding sections, the analysis of the PrPf as either aspect or tense does not do full justice to the status of the construction, and therefore remains somewhat unsatisfactory. In order to remedy this situation and to provide a descriptively more adequate account, radically different proposals have been made as to how to grasp the janus-like nature of the PrPf and English perfects in general.

These approaches gained currency mainly between the 1960s and the 1990s and employ various labels, such as "perfect as phase" (Joos 1964; Markus 1977; Schopf 1984; Meyer 1992, 1995), "perfect as status" (Bauer 1970), "perfect as perfect" (Vermant 1983; Kortmann 1995) or "perfect as orientation" (Kortmann 1991).[90] While some of the accounts (e.g. Vermant 1983) remain relatively vague in their explanations with respect to why the perfect should be a separate grammatical category, others (e.g. Bauer 1970; Markus 1977 or Meyer 1992) see particular semantic characteristics (such as describing an accomplished fact, resulta-

---

sion of terminology and psycholinguistic aspects related to these concepts, see Fenn (1987: 141–144).

[87] For Declerck, comparable to the approach by Huddleston & Pullum (2002: 143), pragmatic speaker focus is the main determining factor between the location of a situation in the past or in the pre-present (Declerck 2006: 110; cf. Section 3.3.3). Note in this context that the different time zones in Declerck's tense system do not say anything about the absolute temporal distance with regard to the moment of utterance (Depraetere 1996: 56).

[88] Others, however, have emphasized that this putatively singular temporal meaning of perfects is "elusive" (Matthews 1987: 111).

[89] The terminology seems slightly odd, as Declerck's "absolute tenses" are in fact relative to the moment of utterance. Furthermore, others doubt that time zones can be defined objectively (Vasudeva 1985: 188).

[90] For a discussion of terminology see also Fenn (1987: 249), who rejects a number of the terms but refrains from proposing alternatives.

tiveness, or current relevance) as sufficient criteria for distinguishing it from the established grammatical categories. Kortmann (1991, 1995) argues *ex negativo* that the perfect does not fully meet the criteria of either aspect or tense (see the preceding sections) and, as a logical consequence, must constitute a grammatical category of its own that is on a structural par, and interacts with tense and aspect (Kortmann 1991: 18 and Figure 3.2.1).

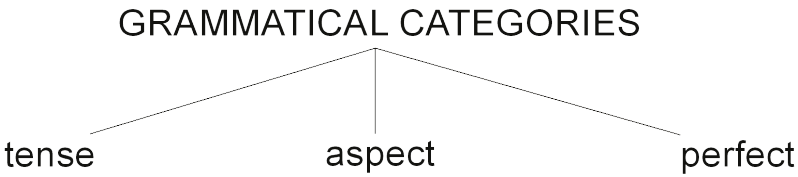


Figure 3.2.1. Grammatical categories as represented in accounts that posit the perfect as a grammatical category of its own

He presumes that tense is deictic and aspect non-deictic. In addition, he argues that situation-internal time can be expressed by aspect and Aktionsart only. Therefore, perfects cannot fall into the aspect category. Like tense, they express temporal anteriority and thus relate to situation-external time. Nevertheless, they do not qualify as tense either, as the sole expression of anteriority is non-deictic. In other words, they "relate [...] some situation to a succeeding reference time" (Kortmann 1991: 20) and not to the coding time, the prerequisite for tenses.[91] Thus, he concludes that a further category, possibly labeled "orientation" (1991: 24), is needed to describe the grammatical status of the perfect (1991: 15–21).

All of these models have a common advantage in that they avoid the inconsistencies and workarounds inherent in the classical "perfect as aspect" or "perfect as tense" approaches (see e.g. figures 3.1.1 and 3.1.2), and are relatively elegant as a result. However, and in spite of repeated expressions of appreciation of their descriptive usefulness (e.g. by Hackert 2008: 129), they have not found general acceptance, as can also be deduced from their absence from reference grammars.

[91] This line of argumentation is echoed by Nordlander (1997: 110; 1998: 13).

3.2.4 Alternative views

In contrast to the approaches described in the previous sections, some theories on the status of the PrPf refrain from a one-dimensional analysis due to the inherently hybrid nature of the construction. As shown above, all of the "perfect as aspect" and "perfect as a category of its own" accounts recognize certain temporal characteristics of the PrPf, but also concede that these may not adequately represent its functional range.

This unease is expressed, for example, in the findings that the PrPf is "*closer* to tense than to aspect" (Wright 1986: 70; emphasis added) or "seems *less tense-like* than the past because it adds a specific kind of focus. [...] The past tense only locates events and is neutral with regard to how they should be seen or conceived" (Petersen 2004: 116; emphasis added). This implies that the PrPf shares properties of tense (but less so than the SPst, for example), and aspect (unlike the SPst). Somewhat inconsistently, however, Petersen adds that the SPst, which is traditionally seen as a tense, also contains features of aspect, though on a different level of analysis that could be labeled "imperfectivity". He comes to the overall conclusion that all verbal forms in English are without exception "non-monadic" (Petersen 2004: 116), meaning they unite features of tense and aspect, and therefore the category "tense-aspect"[92] should be introduced.[93]

Another alternative view can be found in the typological literature and in further analyses of the PrPf in English (e.g. Kallen 1989; cf. Wynne 2000: 174). There, it is recognized that tense and aspect "do not

[92] Cf. Schwenter's category "tense/aspect" (1994: 995) and Thieroff & Budde's suggested category "tense-aspect" with the two values "perf" and "non-perf" (1995: 53), which is on a structural par with tense and aspect. Although employing similar terminology, the latter approach resembles the "perfect as a category of its own" view (see Section 3.2.3).

[93] Wynne takes a similar approach, maintaining that "verbal structures can never be regarded in isolation [but] must be seen in the context of the interplay between their aspectual and semantic values" (Wynne 2000: 174). See also Bache, who considers the attempts at categorizing the PrPf as belonging to either tense or aspect as futile. He argues that, because the PrPf is semantically complex, it can best be described as involving more than one grammatical category. Thus, he proposes a compositional account for the PrPf and suggests the metacategory "situational dependency" (1994: 59), which involves tense, aspect, and Aktionsart (1994: 53–59). See also Section 3.3.4.

always present themselves as separate and neatly delineated categories" (Dahl & Velupillai 2011a). The notion of a hierarchical ordering of different grammatical categories (see sections 3.1 and 3.2) is abandoned in favor of a non-hierarchical building-block model, where tenses, moods, and aspects as individual building-blocks constitute a holistic tense-mood-aspect system of a language (Dahl & Velupillai 2011a) and where a part of this tense-mood-aspect system (e.g. the PrPf) might be realized by a range of surface forms (Kallen 1989: 7).

3.3 Theoretical studies on the semantics and pragmatics of the Present Perfect

Having discussed the different approaches toward the grammatical status of the PrPf in English, this section now turns to a second area of scholarly dispute that has been responsible for a flood of publications and its aim is to present an overview of the positions on the semantic interpretations or "readings" of the PrPf. This is deemed necessary as (i) some of the empirical studies presented subsequently in sections 3.4 and 3.5 rely on terminology and categories established in theoretical works, and (ii) the empirical part of the present study will establish the semantic value of the individual occurrences of the PrPf using a categorization on corpus-based principles. The advantages of this approach, rather than identifying categories by way of introspective reasoning, will become clear in the course of the discussion.

Generally speaking, a division can be made between monosemous accounts, which recognize a single basic meaning or underlying semantic principle that guides the usage of the PrPf, and polysemous accounts, which claim up to seven variable semantic readings of the construction. These readings are either content-sensitive or, in some cases, determined by context or the point of view of the speaker. In addition, compositional theories and views of the PrPf on different levels of analysis within the frameworks of pragmatics and text linguistics are presented. Note that despite the sheer number of accounts, it has repeatedly been observed that many of these accounts are incompatible with each other and, besides the finding that the PrPf may be used in

many different situations (Schlüter 2002a: 29), no real progress has been made in the countless years that scholars have spent establishing a semantic theory of the PrPf (Schlüter 2002a: 60; Sempere-Martinez 2008: 131).[94]

3.3.1 Monosemous accounts

Monosemous accounts of the PrPf share the claim that there is a single general meaning that underlies all other possible semantic interpretations of the construction, although further derived meanings may exist (e.g. due to Aktionsart of the main verb; cf. Portner 2003: 493; Allerton 2008: 27). Thus, approaches are outlined here that see the general basic principle as the expression of either current relevance or anteriority (or posteriority).[95] In addition, the "extended-now" perspective is sketched.[96] All of these approaches criticize polysemous accounts for various reasons. On the one hand, the criticisms rest on formal grounds, specially that the criteria for assigning occurrences of PrPf forms to one category or another, such as temporal adverbials or the Aktionsart of the verbs involved, may be based on our general knowledge of episodes and events and so actually lie outside the PrPf construction or even outside language. Therefore, it is argued that a distinction between different readings of the PrPf is unnecessary from a linguistic perspective as it de-

[94] For older reviews of the literature on the semantics of the PrPf see e.g. Matthews (1987: 120); Brinton (1988: 10–15) or Molsing (2006: 240–244).

[95] A combination of anteriority and current relevance can be found in Biber et al.'s (1999: 460) general definition of the function of the PrPf: "the present perfect is used to refer to a situation that began sometime in the past and continues up to the present". The second part of the argument (continuity up to the present) has been disqualified by some (Kuhn 1989: 537), given that statements such as *I have lived in Chicago but now I live in Montreal* ("existential readings", see Section 3.3.2.1) where there is arguably no actual continuity up the present, are possible. Others, however, take a wider perspective and include logical connections (e.g. living in Montreal could be a consequence of having lived in Chicago for some reason or another) to the present (Allerton 2008: 28).

[96] Note that occasionally also "indefinite" or "resultative" are seen as central the manifestations of the PrPf (Siemund 2004: 419 and the discussion in Depraetere 1998: 597). In the current study, these readings are listed under the polysemous accounts (see Section 3.3.2) as more often than not they are considered as just one particular interpretation among many options.

scribes "nothing but different consequent states" (Moens & Steedman 1988: 20; cf. Moens 1987: 95–103) of the same meaning. On the other hand, it is claimed that categorizations within polysemous approaches are subjective and arbitrary;[97] naturally, however, similar allegations have been made with regard to single meaning theories (e.g. Binnick 1991: 382).

3.3.1.1 Anteriority and posteriority

The broadest possible approach is to see the expression of "anteriority" as the defining feature of the PrPf (Quirk et al. 1985: 4.18). As was pointed out above (Section 2.1), in the typological literature, this is seen as the basic and oldest function of perfects across all languages. The line of argumentation of most advocates of this view is that all other potential readings of the construction may be seen as derived from anteriority and thus constitute merely contextual variants of this central meaning (Bauer 1970: 193–194; Vermant 1983: 112).[98] Some authors leave the type of anteriority they have in mind vague and thus offer no rationale of what differentiates the PrPf from other tenses that can also be used to express anteriority (such as the SPst).[99] Others, who operate within the framework of (neo-)Reichenbachian tense logic, define it more narrowly as anteriority to a point of reference, which may or may not be identical

[97] For a pertinent example see Meyer (1995: 218), who states that "[i]f we are to assume more than one meaning, then in principle we might just as well posit infinitely many because there are infinitely many contexts", although he concedes that a single meaning approach may involve the acceptance of "fuzzy borders" (1995: 201). This view echoes Vermant (1983: 42; see also Portner 2003: 467–475).

[98] It is evident that these authors are only making a statement about the semantics of the verbal PrPf construction as such ("perfect qua perfect"; Bauer 1970: 193) and are disregarding co-occurring elements such as temporal adverbials as well as Aktionsart of the verbs involved. Arguably this approach is too reductionist. As is to be shown in the empirical part of the present study, contextual factors and Aktionsart play an important role in the usage of the PrPf.

[99] See e.g. Bardovi-Harlig (1997) or Canavan (1983). However, the latter concludes that the PrPf as such neither locates the action on a time scale nor expresses tangentiality to the present moment (Canavan 1983: 41). See further Maule, who states that the PrPf is employed when "the speaker has chosen not to signal distance" (1991: 7). Yet, an explanation of the exact nature of this "distance" is lacking.

with the time of speaking (Bauer 1970: 191; Matthews 1987: 167) or, with similar implications, as anteriority to the "time of reckoning" (Harder 1997: 382).

In further accounts, the focus is diametrically opposed as the point of view emerges from some point or state in the past and continues into the post-time of that point (up to the time of speaking) rather than looking backwards from the present. Therefore, in these accounts the expression of posteriority is seen as the main characteristic of the PrPf (Mugler 1988: 86; cf. Vermant 1983: 67).[100]

3.3.1.2 Current relevance

Another approach identifying the semantic properties of the PrPf is the positing of "current relevance" or "present relevance" (Comrie's (1976) label) as the general underlying semantic principle. This view has been widely disseminated but has also attracted repeated criticism for a number of reasons.

In very general terms, the PrPf "indicates the continuing present relevance of a past situation" (Comrie 1976: 52), which, from a diachronic point of view, may be seen as one of the longest-standing implications of perfect forms in general (besides anteriority). However, the exact nature of current relevance might have changed in the course of time, now allowing for the inclusion of more and more different subtypes, such as continuative interpretations (Hantson 2005: 251).[101] The proponents of the notion of current relevance try to provide a comprehensive principle that grasps the dual nature of the PrPf as a grammatical form that is related to the past and present time sphere and creates reference to both (Comrie 1976: 52; 1985: 35; cf. Section 3.2.2 above). In addition, it has been stressed that current relevance excludes the use of definite temporal adverbials, and that the recentness of the situation being re-

[100] See also the definition within Klein's temporal framework, who sees "TT [= topic time; V.W.] in posttime of TSit [= time of situation; V.W.]" (1994: 111; cf. Section 3.2.1) as the general meaning of the PrPf.

[101] See also Michaelis, who states that the subcategorization of current relevance is a matter of "constructional polysemy" rather than "structural ambiguity" (1998: 115–116). For this reason, she favors the current relevance account as opposed to a polysemous approach.

ferred to may be seen as a sufficient criterion for establishing current relevance. However, current relevance does not necessarily imply recentness (Comrie 1976: 60; 1985: 32). Another approach is to state that the PrPf, in contrast to perfective tenses such as the SPst (which describe a situation as discrete), creates various connections to the present (Allerton 2008: 27).[102] Although authors accept that current relevance so defined is a rather vague and fuzzy notion,[103] it is argued that it serves adequately as an overall principle as it allows for the vast number of potential connections and the variety of their individual characteristics (Brinton 1988: 14). It has been suggested elsewhere (Rastall 1999: 83) that, despite its fuzziness, in some cases the current relevance principle is responsible for a reversal of the usual grammatical constraint that the PrPf cannot co-occur with definite temporal adverbials, especially if speakers consider the information they convey particularly newsworthy.[104] This is taken as an explanation of why combinations of the PrPf and definite temporal adverbials occasionally do occur.[105]

Although the vagueness surrounding what current relevance actually represents allows for a fairly inclusive and generic approach, thus making current relevance an appealing concept, this very vagueness has underpinned various criticisms. First, it has been pointed out that current relevance has been defined differently by different authors (Vermant 1983: 21). While this may be undesirable, it is by no means un-

[102] Cf. also Dahl & Hedin, who maintain that current relevance is an important general principle "for the interpretation of event-referring sentences in general" (2000: 399) and in consequence also for the choice between the PrPf and the SPst.

[103] It is indicative in this regard that Veloudis speaks of a "*feeling* of 'current relevance' it [= the PrPf; V.W.] can be *intuitively* associated with" (2003: 388; emphasis added). See also Harder, who states that the PrPf always bears some kind of relevance "if you are not too picky about your criteria" (1997: 384).

[104] However, it should be evident that "newsworthiness" is a very subjective criterion in itself. For further examples of occurrences that apparently violate the constraint see Vermant (1983: 65–84).

[105] Quite the opposite view is taken by Kirsten, however, who argues that the non-compatibility of the PrPf and definite temporal adverbials disproves the applicability of the current relevance principle (1994: 46). This is a line of argumentation that cannot easily be ignored. If current relevance was the general (and potentially overriding) principle, combinations of the PrPf and definite temporal adverbials would be commonly accepted or at least much more frequent (cf. also Comrie's position outlined above).

common in scholarship in general and especially in linguistic treatises of the PrPf. What is more detrimental to the applicability of current relevance as an abstract notion, however, is the subjectivity that is involved on the part of the speakers. In other words, it is not always easy to objectively determine why a speaker wants to express current relevance in one case but not in another, and how it is conceptualized by the individual speaker (Binnick 1991: 382). Second, current relevance has been rejected as being too broad as it is not inherent in the verbal structure of the PrPf construction as such, but rather based on context, for example co-occurring temporal adverbials, general knowledge, or statements in the preceding or following discourse in general. Therefore, it cannot be an inherently semantic principle (Canavan 1983: 40; Mugler 1988: 69–83; Wynne 2000: 168). Third, it has been noted that, due to its vagueness, current relevance could also easily be ascribed to all tenses in one way or another (McGilvray 1991: 48; see also Salkie 1989: 4)[106] as, according to Gricean principles, all meaningful statements should fulfill the relevance criterion (Harder 1997: 382). Fourth, it has been argued in accounts of the "PrPf puzzle" (that is, why the English PrPf, unlike perfects in other languages, does not combine with definite temporal adverbials) that the explanatory power of the current relevance principle is limited as it does not account for why the PrPf does not allow precise localization in time. Therefore, it does not contribute to solving the "PrPf puzzle" and is not generally valid, as it is not operative with, for instance, non-finite forms (Klein 1992: 531–532; 2010: 1241).[107] Fifth, it has been criticized that various layers of verb meaning are subsumed under the umbrella term "current relevance", although in fact they should rather be seen as discrete interpretations on different levels. An onion metaphor illustrates this approach: a particular occurrence of a

[106] Cf. the following (constructed) examples that put current relevance as specific to the PrPf in doubt as the events resulting from these situations are still relevant in one way or another: (i) *Yesterday he broke his arm.* (ii) *Last week he was pulled over by the police and had to hand in his driving license.*

[107] Klein (1992: 531) furthermore recognizes the difficulty of properly defining "relevance". A solution to this problem is proposed by Depraetere, who, in order to preserve the current relevance principle, favors a flexible approach and argues that "the denotation of relevance depends on the different situation types referred to" (1998: 602).

PrPf form may be existential (layer 1) + continuative (layer 2) and potentially also resultative (layer 3). Thus, current relevance again is rejected as being too broad to serve as a semantic principle for the description of the PrPf (Wynne 2000: 168).[108]

In view of this criticism, and although the notion of current relevance holds some explanatory power for PrPf usage, it has been proposed that it represents a concomitant phenomenon rather than the major underlying principle of the PrPf (Kirsten 1994: 47). This view has sparked further discussion as to whether current relevance should be preserved not as a semantic but rather as a pragmatic concept, while others favor a compositional approach (see e.g. Depraetere 1998: 598–602).[109] Without going into terminological detail, note that within these accounts, current relevance has been seen as either inference (Salkie 1989: 6), implicature (Depraetere 1998: 598), or presupposition (Jaszczolt 2009: 86).[110] What these authors have in common is that their approaches allow for current relevance as an overarching principle of the PrPf; however, on a superordinate (pragmatic) level of linguistic analysis.[111] Accordingly, various readings on the subordinate (semantic) level are possible with current relevance as the "common denominator" (Siemund 2004: 414; see also Comrie 1976; Hantson 2005; Davydova 2008).[112]

---

[108] See further Lindstedt, who proposes a slightly different cumulative approach where current relevance combines with a "resultative", "experiential" or "other" meaning (2000: 378).

[109] "Current" in Depraetere's account still represents an inherent semantic notion ("connection between past and present"; 1998: 602).

[110] For Jaszczolt, pastness is the second overarching principle of the PrPf. She also tries to tackle the PrPf "puzzle" and states that definite temporal adverbials do not co-occur with the PrPf "not because they don't convey pastness but because of this common-ground presupposition of current relevance" (2009: 86; but cf. Klein 1992: 531–532; Kirsten 1994: 46).

[111] Further considerations on the pragmatics of the PrPf can be found in Section 3.3.3 below.

[112] This is also why some authors appear twice, for example as proponents of current relevance and a polysemous account, in Table 3.3.1, which gives an overview of semantic theories and categorizations of the PrPf. These authors employ a multiple-layer approach which shows that current relevance and polysemous accounts are by no means irreconcilable. Furthermore, in these multiple-layer approaches, clines of different semantic readings

3.3.1.3 Extended-now (XN)

While "extended-now" is occasionally viewed as one particular semantic reading of the PrPf (e.g. by Davydova 2008) that corresponds to the reading more commonly labeled "continuative" (see Section 3.3.2.4), sometimes it is considered to be the general semantic principle underlying the PrPf. This view, most prominently developed in a treatise by McCoard (1978), derives from the finding that definite temporal adverbials combine with the SPst but not with the PrPf.[113]

It is based on the notions that (i) anteriority is a shared feature of the PrPf and the SPst and therefore cannot be the underlying principle of the PrPf, (ii) the PrPf and the present tense have some shared characteristics, such as reference to the present, and (iii) the main feature of the PrPf is "past-including-the-present" (McCoard 1978: 152), which is responsible for the non-compatibility of the PrPf and definite temporal adverbials as these refer to a past that is separate from the present.[114] The PrPf can therefore be delineated from the SPst as the former expresses "inclusion" (McCoard 1978: 154)[115] but the latter does not.

While the extended-now theory still enjoys some currency (Rathert 2004; Rothstein 2007, 2008; Mittwoch 2008), its value as a general principle for the description of the PrPf has been doubted. This is mainly due to the inherent weakness that it is not able to make meaningful statements about tense choice when temporal adverbials are involved (Dinsmore 1981: 479–480; see also Giorgi & Pianesi 1997: 92). Like the criticisms of current relevance, it has been argued that the extended-now view describes the context rather than the semantics of the verbal PrPf

---

with regard to current relevance, for example from resultative (most current relevance) to experiential (least current relevance) are suggested (e.g. Davydova 2008: 3).

[113] McCoard also provides extensive criticism of current relevance and other PrPf theories. Cf. in that regard Binnick, who states that current relevance theory "is in fact a kind of XN theory, but one in which the strong XN requirement is a *pragmatic* condition" (1991: 272; emphasis original). See also sections 3.3.3 and 3.3.4.

[114] More recent authors advocating an extended-now view have emphasized that this account is not compositional as the definite temporal adverbial constraint is not due to the semantics of the present tense component (= the auxiliary) of the PrPf construction (Rothstein 2008: 175).

[115] In McCoard's hierarchy of grammatical categories, inclusion would be a subcategory of tense, while tense is on the same level as aspect (1978: 152).

construction as such (Mugler 1988: 69–83) and fails to account for non-finite forms of the PrPf (Klein 2010: 1241).

3.3.2 Polysemous accounts

In contrast to monosemous accounts of the PrPf, advocates of polysemous semantic theories postulate up to seven different semantic readings of the PrPf. These readings may vary according to different dimensions and can be seen as interpretations or categorizations of how and when the PrPf may be used. The conceptualization of the semantics of the PrPf is usually broader in the sense that polysemous accounts deliberately include the context in determining the various semantic functions of the construction. In more general terms, the rationale for establishing a polysemous account for tense and aspect forms is the observation that these forms are used in various contexts (Binnick 1991: 104). Attempts at a unified analysis – that is, the definition of a single underlying meaning – are therefore overly abstract. The following sections will describe the various semantic functions individually and provide examples from the corpus data for illustration. At the end, an overview of different polysemous models will be given.

3.3.2.1 Indefinite/existential past

The first semantic function to be outlined, the "indefinite" (Binnick 1991; Leech 2004; Miller 2004a)[116] reading, is almost universally recognized in polysemous accounts of the PrPf (see Table 3.3.1 below). Other, less common labels for this type include "existential" or "experiential" (Mittwoch 2008), "up-to-now" (Declerck 2006), or "remote" (Panitz

[116] Note that other authors use this label differently (e.g. Depraetere 1998; Declerck 2006) to refer to what is commonly labeled the "resultative" category. "Indefinite past" has also been developed as a more general theory that tries to explain the difference between the PrPf and the SPst as due to features outside semantics (e.g. by Vermant 1983), i.e. whether the time that is referred to is definite (SPst) or indefinite (PrPf), although it has been recognized that "[e]ven when indefinite, the perfect has a number of different uses" (Binnick 1991: 98). For a brief contrastive analysis of various perfect theories, see further Binnick (1991: 264–265).

1998). See (1) and (2) for examples of utterances that would typically be categorized as indefinite.

(1) Now here is another picture of the mobile space mobile that I've shown you before (ICE-PHI s2a-035)
(2) Anyway for those of you who have not read Fowles do you think that you will (ICE-IRL s1b-019)

The rationale behind the proposal of the indefinite label is (i) that the focus is on the location of a situation in the past, but with no temporal definiteness (hence, the terminology) and (ii) that, unlike when using the SPst, relevance to the present moment should be conveyed (Bybee et al. 1994: 61–62). Others have recognized that (i) may well also be a criterion for the SPst, particularly if an utterance lacks temporal adverbials (Schlüter 2002a: 178); thus, a "definite" versus "indefinite" dichotomy for establishing the contrast between the SPst and the PrPf is necessarily deficient (Fenn 1987: 162–166). In addition, a number of cues such as typical temporal adverbials (e.g. *ever, never, once*, etc.), expressions involving BE (*Has she been to Madrid?*), prosodic features (stress on HAVE), or more general features such as repeatability have been identified that all strongly suggest an indefinite reading (Mittwoch 2008: 326).

3.3.2.2 Recent past/hot-news

The second semantic function that is seen as a separate category in the majority of polysemous accounts (see Table 3.3.1 below) is the "recent past"[117] or "hot-news" perfect. It is described as expressing recentness and therefore typically occurs at the beginning of news bulletins (see (3)) or statements about recent discoveries, and co-occurs with temporal adverbials such as *recently* or *just* as in (4).

(3) The government has put up a ₱200-million loan fund for the displaced workers. Employee Compensation Commission

[117] Occasionally, the term "recent indefinite past" (e.g. Leech 2004) is employed.

Executive Director Teofilo Hebron said [...] (ICE-PHI w2e-007)

(4) I am appalled that Mitsubishi have sold people a model of car with known defects, and am furious that I have *just* spent $26,000 on a car that performs so badly. (ICE-AUS w1b-029)

Other observations with regard to the recent past reading of the PrPf are that in AmE, recent past contexts more often than not involve the SPst (Miller 2000: 337) and that opinion remains divided over the issue of whether the recent past is merely a subfunction of the "resultative" reading (Michaelis 1994: 127; see also Fenn 1987 and Brinton 1988) or should represent a distinct function of its own (Schwenter 1994: 997).[118]

3.3.2.3 Resultative past

A third semantic function assigned to the PrPf is that of "resultative" or "stative". These labels are chosen as the PrPf may emphasize the change of state and its present consequences (see (5)) due to some kind of causal connection (Portner 2003: 500), while "nil results" (Huddleston & Pullum 2002: 145), as in (6), are possible. Authors acknowledge, however, that a resultative interpretation heavily depends on pragmatics and the surrounding context (Michaelis 1994: 128; Huddleston & Pullum 2002: 145).[119]

(5) Though the days have gone when the French changed their governments with unusual frequency (ICE-NZ w2e-010)

[118] Schwenter concludes that recent past should be seen as a separate reading since what follows a recent past PrPf (e.g. in a news bulletin) focuses on an "elaboration of the past event, not of its present consequences" (1994: 1001). Against this view, it could be argued that the recent past reading for the most part makes a statement about the semantics of the discourse rather than about the function of the construction (and its context) as such (Matthews 1987: 128). Binnick (1991: 99) presents a wholly different approach, seeing recent past readings as mere inference.

[119] See the example *She has been to the bank*, which could imply various notions such as 'she has some money' or 'the cheques are deposited' (Huddleston & Pullum 2002: 145).

(6) I have not become disciple of anybody so far but most of these philosophers had been uh having some or the other impact on me (ICE-IND s2a-028)[120]

Nearly all categorizations of the PrPf include a resultative category and, with regard to frequency, this interpretation is often said to be the most common reading of the PrPf (Bauer 1970: 189; Miller 2000: 324; Leech 2004: 40; Davydova et al. 2011: 303). The prominence of the resultative reading has been attributed to two factors. From a diachronic perspective, resultative contexts are believed to be the oldest contexts in which the PrPf has grammaticalized (see Section 2.4). Additionally, findings from studies on first-language acquisition suggest that the resultative reading is the first one used by children (Slobin 1994: 129). Therefore, the resultative can be seen as a kind of core meaning of the PrPf construction. At the same time, however, proponents of a resultative reading have been criticized for not being able to show how it differs from other interpretations, and for not disproving the claim that it is just a pragmatic addition to the indefinite/experiential reading (Mittwoch 2008: 340). In addition, some authors (e.g. Miller 2000: 330) consider the resultative interpretation redundant as all events in the past (and therefore all the situations described by the PrPf) have consequences in one way or another, and therefore this cannot be a sufficient criterion for establishing a separate reading.

3.3.2.4 Continuative past

The last semantic function that is universally accepted is termed "continuative",[121] while various other labels such as "universal" (McCawley 1983; Binnick 1991; Ritz & Engel 2008), "extended-now" (Miller 2004a; Davydova 2008; Davydova et al. 2011), "inclusive" (Hantson 2005), "persistent situation" (Comrie 1976; Dahl 1985; Dahl & Velupillai 2011c), and the bulky "state-up-to-the-present + habit-in-a-period-leading-up-to-the-present" (Leech 2004) have been suggested. As these designations

[120] Cf. also *She has been to the gas station, but forgot the key to the tank.*

[121] It is also supposed to be a diachronically late development (Boogart 1999: 150).

imply, a situation is viewed as continuing through a temporal period up to and potentially beyond the moment of speech (which is normally identical with the present moment).[122] Further features likely to co-occur with the continuative reading are the expression of general truths (see (7)) and, on a structural level, the use of the progressive (van Rooy 2009: 326) and a restricted set of verbs that are associated with a stative Aktionsart. In the default case, a continuative reading is only possible when an extended time-span is indicated by a temporal adverbial, typically by phrases with *ever*, *since* or *for*, as in (7) and (8). Therefore, empirical studies (e.g. Elsness 1997; Schlüter 2002a; Hundt & Smith 2009) have shown continuative readings to be quantitatively less frequent than other readings.

(7) There there will be even if it's just the sense of change which is so important that uh other areas of the world where they've had decay *for for many years* [...] (ICE-IRL s1b-026)

(8) [...] and the largest, richest chocolates you've *ever* seen! Best cup of coffee I've had *since my arrival in the U.S.* (American coffee is disgusting) [...] (ICE-GB w1b-012)

To complete the picture, the table below provides an overview of polysemous accounts of the PrPf. Note that this table also contains some of the more "exotic" labels for individual readings as well as additional readings (such as "iterative" or "occurrence") that were not outlined in the preceding sections.

[122] Note that Schlüter (2002a: 167) found identity of moment of speech with the Reichenbachian moment of reference in 96% of all cases in his data.

Table 3.3.1. Overview of polysemous semantic accounts of the PrPf (alternative labels in parentheses)[123]

	Readings of the PrPf					Applied by
Two readings	Resultative			Continuative		Bauer 1970, Brinton 1988
	Existential	Universal				Rathert 2004
	Non-continuative			Continuative		Huddleston & Pullum 2002; van Rooy 2009
Three readings	Existential (experiential, indefinite)	Resultative (stative)		Continuative (universal, XN, inclusive, habitual)		Quirk et al. 1985; Fenn 1987; Kirsten 1994; Michaelis 1994, 1998; Kortmann 1995; Tagliamonte 2000; Hantson 2005; Davydova et al. 2011
	Up-to-now	Indefinite		Continuative		Declerck 2006
Four readings	Existential (experiential, indefinite)	Recent past (hot-news)	Resultative (stative)	Continuative (persistent situation, universal, XN)		Comrie 1976; McCawley 1983; Dahl 1985; Binnick 1991; Winford 1993; (Huddleston & Pullum 2002); Portner 2003; Siemund 2004; Miller 2004a; Ritz & Engel 2008; Davydova 2008
	Indefinite	Recent indefinite past	Resultative	State-up-to-the-present + Habit-in-a-period-leading-up-to-the-present		Leech 2004
Five (or more) readings	Experiential/existential/ remote/indefinite	Recent past	Resultative	Persistent situation	Occurrence (*A light has flashed*)	Panitz 1998
	Experiential/existential	Hot-news	Indefinite/ resultative	Continuative	Iterative/repetitive/ declaratory	Depraetere 1998
	Experiential	Recent past	Resultative	Persistent situation/universal	Evidential	Dahl & Velupillai 2011c

3.3.3 Pragmatic approaches

As indicated above, some authors argue that a functional analysis of the PrPf on a purely semantic level is not satisfactory given that (i) cross-categorizations are possible,[124] and (ii) the proposed semantic categories are more or less arbitrary, as demonstrated by the very scope and variety of polysemous accounts that exist (see Table 3.3.1). In addition, it has

[123] See Brinton (1988: 11) and Nishiyama & Koenig (2010: 615) for contrastive tables including further authors. Note that the representation of Declerck's (2006: 312) categorization is not correct in Nishiyama & Koenig (2010).

[124] All the examples mentioned in sections 3.3.2.3 and 3.3.2.4 above, for instance, can arguably be interpreted as either resultative or continuative.

been suggested that the PrPf is pragmatically rather than semantically ambiguous. Operating within a speech act framework, the proposed pragmatic values the PrPf can take (which means these are inferences to be drawn by the addressees) are (i) "persistence", that is, a situation still holds true at the present moment,[125] (iia) "evidential", that is, the truth or likelihood of the object of a speech-act verb is affirmed, (iib) "topic negation", that is, the topic of the further discourse should be determined, and (iii) "commonsense", that is, evidence for a claim is provided. These values are illustrated in (9) to (12) respectively (Nishiyama & Koenig 2010: 633–636; adapted examples).[126]

(9) He has been a member of her household ever since. (inference: 'He is a member of her household')
(10) She has said its losses stand at 6 billion. (inference: 'Its losses stand at 6 billion')
(11) Have you done a lot of camping recently? (inference: 'I want to talk about camping')
(12) You can go around the world in 80 channels. I have sat down on my couch and watched among other things... (inference: 'It is possible to go around the world in 80 channels')

In contrast to this view of the PrPf (monosemous on the semantic level, polysemous on the pragmatic level), others (e.g. Siemund 2004; Hantson 2005; Davydova 2008) take a diametrically opposed position, suggesting that a dual-layer approach is descriptively most adequate.[127] Within the

[125] It has been noted that *since*-constructions, too, always create an existential presupposition of the event (Iatridou 2003: 135).
[126] Note that Nishiyama & Koenig (2010: 637–640) furthermore postulate that one of the main functions of the PrPf is to establish discourse coherence. This position is similar to the textlinguistic approach by Nordlander, who maintains that "in order to understand the intentions of the speaker properly, the hearer needs some additional clarifying means to avoid having to resort to mere guess-work. This clarifying means is provided by the present perfect construction in the form of a cohesive link between the proposition expressed in the present perfect clause and one particular contextual – implicit or explicit – feature which will help the hearer to understand the sentence in a coherent way, namely the temporal point of reference" (1998: 12).
[127] This view can be traced to Comrie (1976).

latter position, concepts such as current relevance are considered as existing on the pragmatic level, while different readings are possible on the semantic level. By the same token, this provides a basis for resolving two areas of dispute, namely variable contexts between the SPst and the PrPf as well as the definite temporal adverbial constraint.

With regard to variable contexts, current relevance is seen as a general notion that expresses the perspective of the speaker,[128] who, unlike with the SPst (Michaelis 1994: 115; 1998: 216–217),[129] wants to convey the relevance of the utterance for the present moment, which is conventionally done by way of the PrPf (Markus 1977: 83; Depraetere 1998: 610).[130] Provided that the addressee is aware of this perspective (Fenn 1987: 197), further possible implications are resultativeness, continuity (Molsing 2006: 248), or "livingness" (Ward 1967: 49) of a past situation.[131] Note that the importance of pragmatics has also been recognized in some genuinely polysemous accounts, which acknowledge that the different readings of the PrPf are context-sensitive and therefore determined by the preceding and following discourse (Panitz 1998: 241).

With regard to the definite temporal adverbial constraint, it has been proposed that it is not part of the semantic structure of the PrPf but rather a pragmatic feature that has grammaticalized over time (Kortmann 1995: 197–198). This could be paraphrased as "the PrPf is not

[128] For an argumentation in favor of the general importance of speaker perspective see Pulgram, who argues that "it is not the event as such but rather the way the speaker wants the hearer to perceive it that determines which tense should be used" (1987: 384; cf. Declerck 2006: 322).

[129] Cf. also Binnick, who states that "the perfect differs only pragmatically, but not in meaning, from the past" (1991: 459).

[130] Note, however, that while Markus speaks of secondary functions of the PrPf, which belong either to the pragmatic or to the semantic level (1977: 83), Depraetere still remains undecided about "whether it is the semantics of the present perfect or a conventional implicature which explains why propositions to now arise whenever an indefinite present perfect sentence is processed" (1998: 611). I suggest that this very much depends on one's definition of current relevance, as for some readings of the PrPf (such as the existential reading), the relevance of the utterance for the present moment can be negated and is therefore arguably not applicable (e.g. *He has worked for a pharmaceuticals company. But this is of no importance for his present job*) (see further Kuhn 1989: 537).

[131] As should be clear from the preceding sections, these are more commonly seen as semantic functions rather than implicatures of the PrPf.

acceptable if the sentence is (explicitly or implicitly) thematically dominated by a past context" (translated from Markus 1977: 83). This is the case where definite temporal adverbials are used, and therefore the PrPf is ruled out. This point will be revisited in the next section.

3.3.4 Compositionality

The preceding sections have repeatedly highlighted the elusiveness of the PrPf, especially with regard to grammatical and semantic categorization. In response to this, some authors have recognized that attempts to comprehend the janiform nature of the PrPf within a single framework will inevitably fall short. Thus, applying Frege's principle, the PrPf has recurrently been described as a grammatical construction that unites features of different (i) morphological forms, (ii) grammatical categories, or (iii) levels of linguistic analysis. Therefore, it warrants the label "compositional". The present section sketches the differing types of compositionality that have been attributed to the PrPf.

On a mere morphological level, it has been recognized that the dual nature of the PrPf is expressed in its form, where the auxiliary HAVE is a present tense form while the past participle represents the past component of the construction (Comrie 1976: 107). This has been interpreted as "form represents meaning" (Hantson 2005: 246);[132] that is, whether the PrPf should be viewed as a present or a past tense (see Section 3.2.2) remains undecided as it unites features of both, namely reference to the present moment + reference to the past.[133] However, this position has not gone unchallenged. Klein & Vater (1998: 218) called it a "gross oversimplification" as it does not properly explain the intricacies

[132] See also Hantson's alternative suggestion that "all that is expressed by the verb *have* is the fact that there is a connection between the subject and the completed action denoted by the past participle" (2005: 249). However, the exact type of connection is left unspecified.

[133] Note that transformational theories employ a similar compositional approach, claiming that a surface structure PrPf has two underlying tenses, namely a matrix sentence that is present and an embedded sentence that is past (Dinsmore 1981: 477–482; McCawley 1983: 104; Fenn 1987: 159–162); hence the term "embedded past theory" for this account. This theory has been viewed as inadequate by later commentators as it does not explain the different uses of the PrPf (existential, resultative, etc.) and, moreover, it is considered untenable from a typological perspective (Binnick 1991: 265).

of why the PrPf is different from other tenses; that is, (i) why variable contexts, for example with the SPst, do exist, (ii) why there are implications such as the definite adverbial constraint, and (iii) why the "present" component is very weak at best in existential readings of the PrPf. Neither is it valid from a typological perspective, as this type of "present + past" compositionality only holds true for English and is not applicable to other languages, in which the "present" component has been fully lost (see Section 2.4), and perfects even have a more extended functional range than pasts. In addition, it can be argued that the PrPf is more than the sum of its parts (present + past), which clearly indicates that it has grammaticalized and should no longer be seen as truly compositional (cf. the discussion for the German *Perfekt* in Rödel 2007: 71–72).

Within approaches that view the PrPf as a mixture of grammatical categories, two or three components are recognized. The underlying assumption is that verb forms are complex and carry more than one grammatical category (Petersen 2004: 116; see also Section 3.2 above), which implies that a categorization as either tense or aspect (or something else) is inadequate and that two components, tense + aspect (Klein & Vater 1998: 225),[134] or even three components, tense + aspect + Aktionsart, constitute the overall meaning of the PrPf, arguably necessitating a broad metacategory such as "situational dependency" (Bache 1994: 59).

Kortmann (1995) provides an alternative approach, with the three components (present) tense + perfect (as grammatical category; see Section 3.2.3) + Aktionsart.[135] He argues that compositionality with regard

[134] These two components can be marked morphologically. HAVE + the past participle (ending) signal perfect aspect, while the morphological ending of HAVE (e.g. the third person *-s*) signals tense (Binnick 1991: 268). This approach would allow for a unified analysis of past, present and future perfects. Nevertheless, it can be argued that it is not fully adequate for English, given that HAVE often occurs without any overt morphological marking and past participles and SPst forms of regular verbs are formally congruent. This would imply that "form represents meaning" in this case is valid with caveats only, and that further factors beyond morphological form have to be taken into consideration when explaining the compositionality of the PrPf.

[135] Aspect (with the values progressive or non-progressive) could be added as a further component. Cf. an alternative account of compositionality of the PrPf by Nordlander (1998: 6) that is based on Aktionsart principles. Suffice it to say that in his approach, the composi-

to the semantics of the PrPf is an appealing concept as it is "intuitive" (1995: 183), mostly operates on binary distinctions (present vs. past, perfect vs. non-perfect, etc.), and succeeds in deriving the various semantic functions (three in his case)[136] of the PrPf from its components (see Table 3.3.2), although he concedes that contextual elements such as temporal adverbials or further lexical items may be relevant (Kortmann 1995: 196).

Table 3.3.2. Semantic compositionality of the PrPf according to Kortmann (1995)

	semantic function		
component	**resultative**	**experiential**	**continuative**
Present tense	Yes	Yes	Yes
Perfect (anteriority)	Yes	Yes	Yes
Aktionsart	Stative	(Telic) events	Not specified
Contextual elements	Optional	Optional	Obligatory (temporal adverbials)

Compositional accounts of this kind (see also Naumann 1998: 193) have been criticized, however, especially in light of approaches that see tense and aspect (and perfect) not as clearly separable but rather as closely related and interacting categories (Michaelis 2006: 221). In addition, one potential weakness of Kortmann's (1995) approach lies in the fact that the continuative should constitute a semantic reading of its own alt-

tionality of the PrPf is that a characteristic of each PrPf situation is development from a punctual/telic phase (first component) into a stative phase (second component). In a similar manner Katz advocates stativity as the main semantic property of the PrPf, in the sense that "perfect predicates denote the state that comes into being when an event occurs" (2003: 215).

[136] Kortmann's category "resultative" includes the perfect of recent past.

hough its defining property is to be found in contextual elements alone.[137]

A third type of compositionality views the PrPf as having a unified meaning consisting of a semantic as well as a pragmatic component. This approach comes in the shape of a theory developed by Klein (1992), who aims to account for the temporal adverbial constraint of the PrPf.[138] He states that if "the position of a time span in relation to TU [= time of utterance; V.W.]" is specified either by lexical means (e.g. temporal adverbials) or by the grammar (e.g. tense forms such as the present tense) it is "p-definite [that is,] it fixes a definite position on the time axis" (Klein 1992: 544). Within his temporal framework, which recognizes two time spans for the PrPf, TT (topic time) and TSit (time of the situation; see Section 3.2.1 above), Klein proposes that a pragmatic "p-definiteness constraint" is in operation. This constraint establishes that either TT or TSit can be expressed by a p-definite item, but not both time spans at the same time. To explain the ungrammaticality of expressions such as (13) he comes to the conclusion that for the PrPf, the expression that is p-definite is the present tense morpheme HAVE, which, unlike other tenses, specifies TT, and therefore the occurrence of any further p-definite expression that specifies TSit, such as definite temporal adverbials, is ruled out (Klein 1992: 543–547).[139] This would otherwise result in a

[137] Kortmann includes "continuative" in his model although he acknowledges that it is not a "central use" of the PrPf (1995: 196). For reasons of consistency and elegance, "continuative" should be left out, which, in consequence, would weaken the claim of comprehensiveness of Kortmann's compositional model. Note that in the vast majority of accounts, however, "continuative" is included among the typical readings (see Table 3.3.1) and that corpus-based analyses have shown that continuatives represent up to 25% of all occurrences of the PrPf (Schlüter 2002a: 160). Besides, it has been claimed that continuative contexts are a significant factor in triggering PrPf usage in BrE (Davydova 2011: 153–154) as well as in L2 and learner varieties of English (Davydova 2011: passim).

[138] Note that Kortmann (1995) draws attention to the fact that comparable accounts had existed before (e.g. Stump 1985, who views the temporal adverbial constraint as a conversational implicature that has grammaticalized) and acknowledges that a comprehensive analysis of the PrPf has to include both (compositional) semantics and pragmatics (Kortmann 1995: 198). In a later publication, Klein develops a radically different compositional account for tense, aspect, and Aktionsart forms that relies on the argument-time structure of verbs, which cannot be explained in detail here (see Klein 2010).

[139] In a criticism of Klein's conceptualization, Rothstein argues that the possibility of only one specifier per sentence is a syntactic rather than a pragmatic constraint (2007: 101; but

pointless qualification of the assertion of an utterance, such as (14), which is comparable to the similarly pointless qualification in (15), where the only p-definite item is the adverbial, while the verb form in the latter (PaPfs) is not p-definite (Klein 2012, p.c.).

(13) *Ted has visited the museum yesterday.
(14) ??Right now, Ted has visited the museum yesterday.
(15) Yesterday at eleven, John had left yesterday at ten.

A more recent compositional theory suggests that not only the temporal adverbial constraint can be explained by pragmatic principles but that, *pace* Kortmann (1995), every feature of the PrPf that goes beyond mere temporality, the truth-conditional semantic component, such as the various functional readings or the notion of current relevance are part of a pragmatic component of the PrPf (Portner 2003: 460).[140] The problem with such accounts is that their pragmatic component is very broad and inclusive by definition, such that the elusiveness of the PrPf is recognized but not explained. Thus, if the pragmatic component is not further specified, its proponents are in danger of creating a "bin category" that comprises everything that cannot be explained as temporal. As temporality (in the sense of past time reference) is by no means exclusive to the PrPf, this would be overly general and inadequate.

3.4 The Present Perfect in non-standard varieties of English

As repeatedly pointed out in the preceding sections, theoretical accounts on the grammatical status and the semantics and pragmatics of the PrPf

cf. Portner 2003: 493). In another paper, Rothstein suggests that the PrPf "puzzle" or "perfect variation", to use his terminology, is explicable not in terms of "some perfect internal component" (2011: 136) but rather due to competition between the PrPf and the PrTs. A different alternative is provided by Boogart, who suggests that, from a historical-typological point of view, the fact that the English PrPf has not developed into a true perfect (see Section 2.1) explains its non-combinability with definite temporal adverbials, as it (still with resultative force) has only a restricted semantic scope (Boogart 1999: 153).

[140] Cf. also Wynne's (2000) "aphragmatic model" of the PrPf, which operates with three layers of meaning, one of them being pragmatic (see Section 3.5.2 for details).

without exception base their observations on standard BrE and AmE and derive their models through introspection with the help of constructed example sentences. Although these accounts are certainly not without their merits, it has been shown that a vast range of theories exists, none of which are generally agreed on. Furthermore, their findings remain incomplete as they do not do full justice to the functional range of the PrPf in non-standard varieties of English, including dialects,[141] pidgins and creoles. Two lines of investigation are worth pursuing: first, the PrPf may be replaced by other surface forms that express functions traditionally ascribed to the PrPf; and second, the PrPf may have an extended or restricted functional range in certain varieties.

It has been suggested that findings from purely theoretical approaches should be enriched by way of contrast with actual occurrences of the phenomenon under investigation in non-standard data, even if this would entail abandoning or at least complicating an elegant (theoretical) model (Anderwald 2009: 194). In addition, an outline of findings on the PrPf in non-standard varieties should help to further contextualize the findings of the empirical part of the present study, as some of the World Englishes under investigation – or at least their spoken versions – may be seen as non-standard varieties themselves; thus, it seems feasible that parallels exist.

With regard to the usage of the PrPf in dialects of BrE and *cum grano salis* also in spontaneous spoken English in general, opinion is divided insofar that some claim that the PrPf in these registers faces strong opposition from the SPst[142] and, when viewed as a resultative, potentially from other constructions, as in *That's another student given up physics* or *Here's the letters written and signed* (Miller 2000: 324, 347–349). A further assertion is that its functional interpretation depends in the

[141] Dialects are conceived of in the more traditional sense, i.e. regional (predominantly rural) dialects of the British Isles and Northern America in contrast to standard BrE and AmE.

[142] Miller suggests that the SPst is taking over more and more ground from the PrPf in all registers apart from written standard BrE (2000: 324; 2004a: 231; cf. also Schottman 1997). This view is echoed by Allerton, who maintains that due to the influence of AmE and non-native Englishes (!), contexts with a current relevance reading are increasingly expressed by the SPst, while continuative contexts are increasingly expressed by the present tense (2008: 28). This hypothesis should be tested in a separate study.

majority of cases on co-occurring adverbials (Miller 2004a: 230).[143] Others maintain that the structures and functions are basically the same as in standard BrE, although variants such as invariable forms for the SPst and past participle (*She has went*),[144] auxiliary omission (*You seen him since you been on?*; Vermant 1983: 65), BE-perfects (*The train is arrived*) or, in North-American dialects, DONE-perfects (*I (have) done let her go*) do occur, albeit rarely (Van Herk 2008: 48–49).

Another variant of perfect-marking that has been observed in non-standard varieties by Poplack & Tagliamonte (2001: 4) and that is believed to be especially salient in English-based creoles is zero marking of tense and aspect, which is consequently also applicable to the PrPf. These authors suggest, however, that creole verbs with zero marking rather resemble the English SPst and its functional interpretations (2001: 114). Further ways of expressing perfect semantics that are recognized in creole studies and that would typically require a PrPf in standard English are lone past participles (*He gone to the beach*),[145] BEEN- (*I*

[143] His finding that temporal adverbials occur proportionally more with instances of the PrPf than with the SPst (Miller 2004b: 322) leads Miller to suggest a development of "anti-grammaticalisation" (2004a: 230), i.e. a dissociation of the general category perfect into different functions depending on the temporal adverbials in the context. This claim may be somewhat overstated, as temporal adverbials in fact do not co-occur with the majority of PrPf constructions (Elsness 1997; Schlüter 2002a; Hundt & Smith 2009) and are therefore an improper trigger for determining the readings of the construction or even for the constitution of different constructions as suggested by Miller. A related statement, namely that "the classic adverb-less Perfect exists only in formal written English" (Miller 2004a: 231), would imply that the frequency of temporal adverbials in spoken and informal registers should be higher than in written and formal genres (see Section 6.4 below).

[144] Trudgill & Hannah have a point in stating that in standard English the formal distinction between the SPst and the PrPf by way of distinct SPst and past participle forms + additional marking with HAVE is redundant, which explains why many dialects can do without it (2008: 3). Note that also the morphological form of the participle can be non-standard, e.g. through "weakification" (Anderwald 2009: 183)/regularization (*blowed, stealed*) or substitution of a weak for a strong form (*fit* for *fought*, *riz* for *rose*) (Poplack & Tagliamonte 2001: 152–156).

[145] Formally, lone past participles and SPst forms are congruent in some cases, so it might be argued that the SPst is a variant of the PrPf (see Winford 1993: 162–163 for a pertinent example). Miller (2004a: 233) opposes this view as the SPst pressurizes the PrPf and therefore should not be seen as a variant of it but rather as a competing construction. Note that distinguishing between a formally congruent past participle and a "proper" SPst form may be nearly impossible in natural data as other factors could come into play, such as phono-

been watched you for a week) or DONE-perfects (see above),[146] BE-perfects (see above), verb stems (*I never like the city*), *ain't* + past participle/SPst (*Me ain't met him yet/He ain't wrote yet*) and three-verb clusters with auxiliary BE (*I'm done been over there plenty*) (Tagliamonte 2000: 331–332). It has repeatedly been argued that these different formal variants correspond to the different types of meaning that have been established for the PrPf in a more or less systematic way (Hackert 2008: 139; see further Harris 1984 and Winford 1993), such that, for example, continuatives are most likely expressed by zero marking, resultatives by BEEN-constructions, etc. In light of the variability of the sematic models in circulation, it should be evident that further work is needed to determine whether these relationships of correspondence adequately mirror linguistic reality (see further Section 6.4).

Yet another noteworthy finding is that, despite the claim in earlier studies that the PrPf is nearly non-existent in creoles, speakers of early African American English as a putative creole-derived variety used PrPf forms, in the vast majority of possible contexts (Van Herk 2008: 54).[147] The claim of absence of the PrPf is mainly due to methodological factors such as the choice of language material on which the earlier analyses were based (Van Herk 2008: 64–65). From this finding it can be deduced that speakers of World Englishes and particularly of L2 varieties also regularly employ the PrPf pattern and use alternative formal variants only rarely, a hypothesis that will be tested in the present study.

With regard to an extended functional range, the use of the PrPf as a kind of narrative tense comparable to the situation in many European languages has been described as possible in both L1 (Hundt & Biewer 2007: 263) and L2 varieties (van Rooy 2009: 311; Davydova 2011: passim). Further, it seems to constitute a relatively new but salient devel-

---

logical reduction (*I Ø set the watch*) and thus the creation of "apparent preterites" (Van Herk 2008: 64; cf. Elsness 1997: 347–348; 2009b: 243).

[146] BEEN- and DONE-constructions can also be interpreted as general past markers (eWAVE Feature 111 2012), which would imply that their primary semantic function is to create past time reference.

[147] He also gives examples of further ways in which reference to the past is created in his corpus of early AAE correspondence, and contrasts his findings with an earlier unpublished study of a creole variety. In his study, the numbers of these variants are negligible relative to the share of occurrences that resemble standard usage.

opment in non-standard AusE (Engel & Ritz 2000; Ritz & Engel 2008; Ritz 2010).[148] The observations that the PrPf is used in contexts that would typically require a SPst and that it combines with certain definite temporal adverbials are interpreted from a functional perspective in such a way that this use of the PrPf both allows the expression of a "retrospective look onto a past situation" (Ritz & Engel 2008: 156) and opens a kind of post-time gap where situations and events that followed this past situation may be located (Ritz & Engel 2008: 156). It is claimed that this often creates a "mirative effect" (Ritz 2010: 3410), that is, expressing wonderment about unusual or unpredictable events.[149] These findings have to be qualified somewhat as this type of PrPf usage seems to be restricted to very particular genres (police reports and after-match football interviews (in BrE); Walker 2008; 2011).[150] However, the conclusions that perfects are a rather unstable category and that this type of perfect shares features of both the PrPf and the SPst, and therefore covers a middle position contrasting with these two tenses (Ritz 2010: 3411–3416), seem valid. It has to be kept in mind, however, that this development appears to constitute an idiosyncratic feature; thus, it remains to be seen whether long-term implications will emerge, such as a potential extension of the functional range of the phenomenon, and whether similar changes will be observed in further varieties of English (see Section 6.6).

[148] The accounts of the narrative PrPf in AusE operate within a current relevance framework, although the authors criticize the vagueness (see Section 3.3.1.2) of the concept (Engel & Ritz 2000: 120).

[149] Similar constructions can also be observed in a number of other languages, e.g. Turkish (Binnick 1991: 99).

[150] Therefore, Walker suggests that "narrative" should be established as additional distinctive semantic reading of the PrPf (cf. the polysemous semantic models above) that has persisted throughout the history of English (2011: 71, 75). Walker labels this the "present stasis hypothesis" (2011: 75); that is, the re-emergence of the earlier phenomenon of interchangeability between the PrPf and the SPst, or, in other words, a non-completion of the functional division between the two forms that is especially salient in non-standard English (2011: 83). Note, however, that he excludes combinations of the PrPf and definite temporal adverbials from his definition of a narrative perfect, stating that the evidence for these combinations is "close to anecdotal" (Walker 2011: 72).

3.5 Corpus-based (cross-)variational studies

In order to complement the suggestions made in theoretical accounts and in surveys of the PrPf in non-standard varieties of English in general, this section outlines the findings of a selection of earlier corpus-based studies. Particular attention is paid to studies in the cross-variational tradition, which the present study seeks to follow. It has to be noted that in comparison with the sheer number of extant accounts on the grammatical status and functional range of the PrPf, the amount of corpus-based work is still relatively scarce, although the last decade saw the publication of both articles and book-length treatises of the subject.

As most of the newer studies (see also Schlüter 2006) contain more or less extensive reviews of previous empirical work (such as Dubois 1972; Vermant 1983 or Meyer 1992; 1995), only selected publications are discussed here. Studies were included if (i) they can be seen as pioneering from a methodological angle and have served as both a point of reference for the findings of later work and as important starting points for discussion (Elsness 1997), or (ii) they have been neglected by the wider research community as they have remained unpublished (Wynne 2000) or were published in a language other than English (Schlüter 2002a),[151] or (iii) they are very recent (Hundt & Biewer 2007; van Rooy 2009; Davydova 2011).[152]

[151] Mukherjee's hope that Schlüter (2002a) "will find an international target audience in spite of its being published in German" (2005: 203) has obviously not been fulfilled. It is indicative that Schlüter (2002a), despite its topicality, is absent even from the references of works such as Davydova (2008; 2011); Davydova et al. (2011) or Hundt & Smith (2009), although it has to be conceded that a follow-up meta-study in a smaller format (Schlüter 2006) that includes data from Schlüter (2002a) is occasionally recognized (e.g. by Hundt & Smith 2009).

[152] Two relevant recent contributions apparently missing from the current section (Yao & Collins 2012; Seoane & Suárez-Gómez 2013) are directly compared to the findings of the present study, and are therefore introduced below.

3.5.1 Elsness 1997

Elsness (1997) conducted a comprehensive study focusing on the PrPf/SPst alternation[153] and contrasting PrPf usage in BrE and AmE. He also provides a diachronic perspective and discusses a number of semantic PrPf theories. By way of quantitative analysis of corpus data[154] in combination with elicitation tests from speakers of BrE and AmE, Elsness arrives at various conclusions with regard to the PrPf. First, it is used when a situation is "located within a time span not clearly separate from the deictic zero-point" (1997: 348), that is, the moment of utterance. Therefore, temporal adverbials that refer to a past time zone wholly distinct from the deictic zero-point do not occur with the PrPf, while a certain group of adverbials freely combines with both the PrPf and the SPst (1997: 231) and some (*already* and *yet*) that typically occur with the PrPf in BrE are predominantly used with the SPst in AmE. Elsness also finds that the PrPf correlates with negation. He rejects current relevance as a concept contrasting the PrPf from the SPst, claiming that other factors such as the presence or absence of temporal "anchors" have a stronger influence on the choice, although the PrPf may carry a current relevance connotation (1997: 222–223). He also observes that the PrPf is the less frequent and thus more marked member of the SPst/PrPf paradigm, but that the two forms are variants in certain contexts. With regard to variational differences he finds that the PrPf is more frequent in BrE than in AmE (1997: 229–230), and that the text type/mode of discourse and Aktionsart of the verbs involved are also important factors in determining frequency; for instance, the PrPf is more frequent in texts

[153] Due to this focus of his research, he categorizes the PrPf as a tense. At the same time, he recognizes that this view is problematic with regard to Past Perfect and non-finite forms, which are non-deictic (Elsness 1997: 349).

[154] Elsness uses the tagged versions of the LOB and BROWN corpora (c. 1 million words of written data from 1961 for BrE and AmE each), which he fuses with spoken data of the Survey of English Usage (BrE; no word count provided) to create the untagged CONTCORP corpus (Elsness 1997: 79–96). For the diachronic analysis, he compiled the untagged HISTCORP corpus, which comprises material from the 9th until the end of the 18th century (1997: 254–256). Note that, instead of absolute word counts, counts of the "recorded verb forms" are provided in the appendices that shed light on the composition of the corpus material.

with a present-time orientation, such as informational texts, and when achievement verbs are used (1997: 234–235). Finally, from a diachronic point of view, his analysis – which is generally in line with other accounts on the history of the PrPf (see Section 2.4) – suggests that PrPf usage increased from OE to 1800 and has declined thereafter in both BrE and AmE. This leads him to conclude that "this verb form is now losing ground" (1997: 341), which may eventually continue into a further decline of PrPf frequencies in both BrE and AmE (1997: 348).

Elsness has been credited for the richness of the illustrative data he provides, his finding that contextual factors play a larger role in determining the choice between the SPst and the PrPf than had previously been estimated, and his critical discussion of theoretical accounts of the PrPf (e.g. within the framework of formal tense logic) as well as the concepts of current relevance and resultativeness in light of his corpus data (Brinton 1999: 358–359). However, some have suggested that his finding that the PrPf is decreasing in frequency in both varieties has to be put into perspective in view of grammaticalization theory (Tagliamonte 1999: 377). Other authors have highlighted some further points from which his study would have profited. Some question the reliability of the numbers provided (Brinton 1999: 358; Schlüter 2002a: 58), but his mode of presentation is most likely merely due to the technical constraints Elsness had to cope with at the time. Further criticisms, such as using unbalanced corpus material (Brinton 1999: 358) or presenting findings that are based on "logical conclusion[s]" (Wynne 2000: 39–40) rather than actual corpus data are of a methodological nature. Yet, as mentioned above, this study can be seen as a pioneering attempt at a comprehensive analysis of the PrPf in the two major varieties of English.

3.5.2 Wynne 2000

The motivations behind Wynne's study are manifold. He reviews earlier theories and empirical findings on the PrPf in AmE and BrE in light of corpus data and examines the frequency of temporal specifiers with a particular focus on the implications of teaching these "signal words" in language pedagogy. As he finds previous semantic theories of the PrPf

unsatisfactory, he develops an "(a)phragmatic" model that is based on different levels of linguistic analysis with more general (e.g. temporal) and more specific (e.g. aspectual and pragmatic) meanings that combine to constitute the overall meaning of the construction. In addition, he also tries to identify factors that determine a speaker's choice between the PrPf and the SPst.

In comparison to earlier studies, Wynne (2000) uses a wide selection of synchronic corpus material, totaling c. 5.5 million words[155] and consisting of both spoken and written BrE and AmE data.[156] He uses automatically tagged versions of these corpora,[157] enabling him to search in a more sophisticated manner than e.g. Elsness (1997). Based on these data, he arrives at a number of results. First, *pace* Elsness (1997), he rejects the hypothesis that the PrPf is losing ground to the SPst, and adds a note of caution that more occurrences of the SPst do not necessarily result in fewer occurrences of the PrPf (2000: 45). This suggests (i) that merely comparing absolute numbers of the PrPf and the SPst is not advisable and that relative numbers should be considered instead (see Hundt & Smith 2009: 51 for another example of the latter approach) and, furthermore, (ii) that studies contrasting PrPf and SPst usage should only do so in strictly variable contexts. Second, Wynne's data provide evidence that, overall, the PrPf is in fact slightly more common in AmE than in BrE (2000: 46), which is striking in light of findings

[155] Wynne includes additional material for his analysis of temporal specifiers, bringing his total word count up to c. 13.5 million words overall (2000: 21).

[156] His data comprise the *Corpus of Spoken Professional American-English* (CSPA; c. 2 million words of spoken AmE data from 1994 to 1998; topics: academics and politics), the UKSPOKEN subcorpus from the *Cobuild-Collins Bank of English* (COB; c. 10 million words of spoken BrE data from 1991 to 1999; topics: academic lectures, seminars and presentations, radio phone-ins), the *TIMES* corpus (c. 2 million words of written BrE data from 1993; topic: domestic news) and the *WASHINGTON POST* corpus (c. 350,000 words of written AmE data from 1998; topics: domestic and political news). For the exact layout of the material see Wynne (2000: 21–26).

[157] A reliability test of his tagger yielded a recall of 99.5% and a precision rate of 89.2% (Wynne 2000: 32). It has to be noted, however, that the PPT tagging software used has somewhat less power in comparison with other POS taggers such as CLAWS, as it is only able to identify instances of HAVE and its negative forms directly followed by a past participle (Wynne 2000: 28–31). All instances with intervening material (*I've basically just said that...*) are therefore excluded from Wynne's analysis.

from other cross-variational corpus-based studies (Meyer 1995: 206; Elsness 1997: 229–230; Hundt & Smith 2009: 48). Third, with regard to temporal adverbials, Wynne's data suggest that the majority (67.6%) of all PrPfs occurs without a temporal adverbial (2000: 124), and that for most temporal adverbials (apart from *just*, which occurs more often with the SPst in AmE), usage and correlation values for temporal adverbials and the PrPf in BrE and AmE are similar. This implies that preconceived notions of temporal adverbials that are supposedly more characteristic of either one or the other variety should be revised (2000: 111).[158]

The largest portion of Wynne's work, however, is dedicated to the development of a semantic model of the PrPf. After criticizing various other PrPf theories, such as indefinite past or current relevance (2000: 114–168), he discusses the issue of the grammatical categorization of the PrPf and concludes that "the distinction between tense and aspect is arbitrary and unwarranted" (2000: 176), as the full meaning of the construction only emerges from the interplay between the two categories in addition to other contextual and non-linguistic elements. As a solution, he proposes a model of the PrPf that operates with three layers of meaning: (i) the basic meaning of the PrPf is the expression of anteriority and an open event ("aphragmatic") time frame, (ii) on the second level, aspectual characteristics (Aktionsart, temporal specification, progressive) come into play, which allow for different semantic readings (accomplishment, achievement, etc.), and (iii) the third level is pragmatic, which allows for resultativeness, recentness or current relevance, while pragmatic considerations are also responsible for the choice between the SPst and the PrPf (2000: 218; cf. also Section 3.3.3 above). At the same time, however, Wynne recognizes three basic readings of the PrPf, which he labels "finished event"/"iterated finished event" and "event which has started in the past and is either in progress or interrupted at MOU [= moment of utterance; V.W.]" (2000: 207), which roughly corre-

[158] As a large number of temporal adverbials occur with both the PrPf and the SPst, from a pedagogical point of view he doubts the usefulness of a number of "signal words" that are commonly mentioned as being typically used with the PrPf (e.g. *ever, recently, just*). Instead, he suggests that certain other expressions (e.g. *in/over the last/past* besides the commonly recognized *since, so far, yet*) co-occur more regularly with the PrPf in his data (Wynne 2000: 111–112).

spond to the labels "non-continuative" and "continuative" (Schlüter 2002a).

In contrast to other treatises, Wynne's (2000) approach has to be credited for attempting to reconcile many earlier findings on the PrPf within a single model, even though this implies that the model is not very elegant. This lack of elegance and the fact that the model represents a fusion of previously established theories rather than a radically new and innovative approach may explain why it has not found its way into mainstream PrPf theory.[159] It has to be added that, despite the title of the study, the aphragmatic model is not corpus-based but derived through introspection, with the corpus data being used merely for illustrative purposes.

In addition, various methodological criticisms of his corpus study could be made. The selection of the different corpora to contrast BrE and AmE usage does not seem to be felicitous either quantitatively or qualitatively, due to their differing sizes (e.g. 2 million words of written BrE versus 350,000 words of written AmE) and in terms of content (business meetings and press conferences for AmE versus radio phone-ins and academic teaching for BrE). That this choice produced some odd results (e.g. a higher frequency of the PrPf in AmE than in BrE) is becomes clear when comparing the findings of other studies with a similar focus. Furthermore, potential register effects are considered only marginally, again mainly due to the fact that the corpora available to Wynne did not have directly comparable subcorpora.[160]

3.5.3 Schlüter 2002a

Although the motivation behind Schlüter's (2002a) study (see also Schlüter 2000 and 2002b) is primarily of a pedagogical nature (i.e. to establish a didactical, corpus-based grammar that may lay the groundwork for teaching materials on the PrPf for learners of English), he pro-

[159] But cf. Liszka, who states that, from a language-acquisition point of view, a multi-layer model is adequate as morphosyntactic form, semantic features, and pragmatics have to be learned to use the PrPf correctly (2003: 16). This view echoes Bardovi-Harlig (1997: 382).

[160] Wynne is aware of some of these limitations, e.g. that his corpus data are not representative, but states that his results "will be indications of a certain tendency" (2000: 28).

vides one of the most comprehensive empirical analyses of the PrPf to date. Besides an in-depth critical review of earlier theoretical and empirical studies, he offers a wealth of frequency tables and graphs that can be employed for the purposes of comparison. He also develops a model for the semantic description of the PrPf that is based on the data from the corpora he uses. In addition, Schlüter highlights the advantages of using tagged corpora, sketches areas of further study, and explores the importance and frequencies of temporal specifications and co-occurring verbs.

Schlüter uses various corpora of BrE and AmE,[161] and reveals a number of insights as regards PrPf usage. First, in the area of register and genre differences, he finds an overall higher frequency of PrPf constructions in spoken language and in non-fiction texts in contrast to fictional texts. Furthermore, his data reveal a higher frequency of the PrPf in BrE. For written material, however, Schlüter comes to the conclusion that the relative frequencies of PrPf constructions are more similar between texts of the same genre across varieties than between texts from different genres of the same variety (Schlüter 2002a: 109). This indicates that differences in distribution are mainly due to genre effects and less so to regional differences (Schlüter 2002a: 137), and suggests that the two dimensions genre and variety (and text type) merit special attention. In addition, further data with regard to different formal and grammatical variants of the PrPf (simple vs. progressive vs. passive or questions vs. affirmatives vs. negative statements, etc.) are provided, which suggest that the PrPf almost exclusively occurs in the simple form and affirmative statements. Second, he observes that only a minority (c.

[161] His data comprises a sample (CONV; c. 110,000 words) from the *Corpus of English Conversation* (CEC; c. 200,000 words of spontaneous BrE conversations; recorded from 1953 to 1976), which is part of the *London-Lund Corpus* (LLC), and selections from LOB and BROWN (c. 135,000 words from sections A-J for non-fictional and c. 240,000 words from sections K-R for fictional texts for each corpus). These selections are relabeled LOB/BROWN S ('Sachtexte', i.e. non-fiction) and LOB/BROWN F ('fiktionale Texte', i.e. fiction). All in all, Schlüter identifies a total of 2,828 PrPf occurrences (including 691 from spoken texts). Note that no spoken data are available for AmE and that, for the sake of comparison, data from the BNC (three sections of c. 1 million words each) are occasionally employed, too. For further information on the composition of the corpus material see Schlüter (2002a: 76).

33%) of all PrPf occurrences are specified by a temporal adverbial, but that less central functions of the PrPf (see below) are more likely to require such a specification. The former finding is in line with other analyses (such as Wynne 2000 or Hundt & Smith 2009), while the latter was reproduced in another dataset (see Section 3.5.5). On a related note, Schlüter argues against the "recent past"/"hot-news" reading of the PrPf as (i) its indication by way of *just*, *recently* or other adverbials denoting recentness is very rare overall (c. 5%) and (ii) recent past implications can be found in all of his PrPf functions.

The main focus in Schlüter (2002a), however, is the development of an alternative model for the description of the semantic functions of the PrPf. The rationale behind this is a discontent with the variability, apparent subjectivity, and inadequacy of earlier approaches (Schlüter 2002a: 60; see also Section 3.3). The innovative feature of Schlüter's model is that it is derived from the corpus data[162] and, unlike many other models, not based on mere introspection. To achieve this, he proposes a three-level system of *Geschehenskonzepte* ('concepts of happenings') with binary values for each of the levels continuity (continuous vs. non-continuous), multiplicity (simple vs. multiple), and status (dynamic vs. stative),[163] and categorizes all occurrences of the PrPf in his data according to these basic semantic values. Schlüter emphasizes the importance of continuativeness and therefore distinguishes, on the first level, between two basic functions of the PrPf, namely "indefinite past", that is, events that started at an indefinite point in the past and are fin-

[162] Note that for the development of the semantic model, only c. 50% of all occurrences (= 1,418) of the PrPf in the corpus material were used (for details see Schlüter 2002a: 148–150).

[163] These are based on the "situation types" as outlined by Quirk et al. (1985: 4.27ff) and represent basic semantic variables according to Schlüter (2002a: 61). In principle, these constitute $2^3 = 8$ potential semantic readings. However, only four of these are actually realized (Schlüter 2002a: 156). In Chapter 7 of Schlüter's analysis the *Geschehenskonzepte* are further elaborated into *Lesarten* ('readings') labeled "process", "accomplishment", "transitional event", etc. (see Schlüter 2002a: 202 for an overview) in order to do justice to the semantic variability of the PrPf. This approach leads to a meticulous categorization of individual occurrences of the PrPf.

ished at the moment of utterance,[164] versus "continuative past", that is, events that have started in the past and continue up to the moment of utterance or terminate only shortly before it.[165] On the next level (multiplicity), the basic function "indefinite past" can be further subdivided into single versus multiple acts and events, while "continuative past" only occurs with single events. On the third level (status), "indefinite past" exclusively occurs with dynamic status, while "continuatives" occur with both possible values. In conclusion, Schlüter proposes four principal semantic readings of the PrPf: (i) indefinite – single act/event, (ii) indefinite – multiple act/event, (iii) continuative – continuous act/event, and (iv) continuative – state (see Section 4.3.4 for examples). It is worth noting that the frequency distribution between the readings is not equal. Indefinites account for more than 80% of all occurrences[166] and are more prominent in non-fiction texts, while regional differences seem to play only a minor role. Therefore, Schlüter argues that the main function of the PrPf is reference to an act or event that has taken place some-

[164] Schlüter follows Reichenbach (1947) in recognizing a distinction between the point of reference and the point of speech/moment of utterance. However, he finds that, as a rule, these two points coincide (Schlüter 2002a: 172).

[165] Note that this view of continuativeness that allows for termination shortly before the moment of utterance is markedly different from other definitions of the concept (Schlüter 2002a: 200). Cf. the description of the PrPf by Declerck (2006), which bears some striking similarities to Schlüter's model and also recognizes an "indefinite" and a "continuative" reading besides an additional "up-to-now" reading (Declerck 2006: 308). Not surprisingly, earlier versions of Declerck's model of a tense system also receive Schlüter's approval (2002: 45). Note that a later empirical survey using a variable rule analysis corroborated the importance of indefinite contexts for the PrPf but, *pace* Schlüter (2002a), claimed a strong favoring effect of recent past/hot-news contexts (Van Herk 2010: 55–57).

[166] The exact breakdown is 62.8% for indefinites – single act/event, 19.5% for indefinites – multiple act/event, 8.9% for continuatives – continuous act/event and 8.8% for continuatives – state (Schlüter 2002a: 295). Note that Schlüter's model corresponds strikingly to an earlier one proposed by Vermant, who recognizes anteriority to the moment of speech as the basic meaning of the PrPf. He then identifies four "'class meaning[s]' of single-event, iterativity, continuation or ambiguity" and comes to the conclusion that "additional shades of meaning such as recency, resultative, explanatory" (1983: 42) are determined merely by contextual factors. One weakness of Vermant's approach is the "ambiguous perfect" that "occurs in sentences in which it is impossible to say whether the event occurred once, repeatedly or continuously up to now" (Vermant 1983: 64). Particularly when no temporal specification is present, as is the case with the vast majority of PrPf occurrences, this category is potentially too broad to be meaningful.

time in the (indefinite) past and that is finished at the moment of utterance. Continuative readings, in contrast, are more marginal (Schlüter 2002a: 150–171). The salience of the indefinite function is in line with findings from a number of earlier works that are examined in a subsequent meta-analysis by the same author (Schlüter 2006: esp. 146).

It is evident from the preceding description that Schlüter (2002a) has to be credited for his original approach toward the semantic description of the PrPf.[167] He convincingly exposes the weaknesses of traditional approaches toward the categorization of the PrPf and develops an alternative theory that is strictly based on empirical findings and can be applied to other corpus material. The main difference with earlier (and also with many later) corpus studies is that Schlüter's categories are deduced in retrospect rather than categorizing the instances of the PrPf according to pre-established labels arrived at by introspection. In the present study, for reasons of comparability, these categories are preferred over those of the introspective models. It has been pointed out that Schlüter's study occasionally lacks a connection to the wider linguistic context, potentially due to the vast amount of data the author tackles (Mukherjee 2005: 202). It is evident that Schlüter's conclusions are based on relatively old corpus material (the most recent data are from 1976) and that due to the nature of the data only a very rough split into different genres is possible, which somewhat limits the overall explanatory power of the survey, in particular for the spoken register. Furthermore, non-standard usages of the PrPf (for example combinations with definite temporal adverbials) are not systematically included. Nevertheless, the many quantitative findings provide a useful point of reference for further studies of the PrPf as well as room for extension.

3.5.4 Hundt & Biewer 2007

This contribution can be seen in the tradition of investigations of the variation between the PrPf and the SPst (such as Elsness 1997 or Wynne 2000), but is methodologically different from most other corpus-based

[167] The transferability of his results to and the relevance of corpus studies for the domain of language pedagogy have also been acknowledged (Mukherjee 2005).

studies as it relies exclusively on data from online newspapers. Its main focus is to test the hypothesis that recently codified native varieties such as AusE and NZE as new "centre[s] of gravity" (Hundt & Biewer 2007: 249) influence other, geographically close, varieties of English, such as PhiE, SinE or FijiE. In other words, the authors aimed to determine whether the L2 varieties under investigation are (still) oriented toward one of the "old" standards (e.g. BrE for FijiE and SinE or AmE for PhiE) or rather toward the "newer" ones (AusE or NZE). The domain of grammar seems particularly apt to test such a potential orientation, as it is usually assumed that grammatical items are less associated with marking for regional identity in contrast to lexical or phonological features (Hundt & Biewer 2007: 250). At the same time, the authors explore the strengths and weaknesses of relying solely on web-derived data.[168]

First of all, contrary to earlier studies (Quirk et al. 1985: 4.22; Meyer 1995: 206; Biber et al. 1999: 462–463), Hundt & Biewer find that the PrPf is not less frequent in AmE compared to BrE. In addition, no clear distinction emerges between varieties traditionally oriented toward either BrE or AmE, but rather between L1 and L2 varieties, with the L1 varieties being more PrPf-friendly. Hundt & Biewer speculate that this split between L1 and L2 varieties is mainly due to substrate influence (2007: 255–257; but cf. Davydova 2011). Secondly, with regard to the co-occurrence patterns of the PrPf and the SPst and the adverbials *already*, *yet* and *just*, a mixed picture emerges that does not allow for firm conclusions, as all adverbials appear mainly with the PrPf, but also with the SPst across all varieties (apart from the combination *yet* + SPst in FijiE). Occasional instances of PrPf usage in typical SPst contexts do occur, though rarely (2007: 259–262). On the basis of these findings, Hundt & Biewer conclude that a growing influence of transplanted native varieties such as AusE and NZE on geographically close L2 varieties cannot be observed in their data, but that further corpus-based work is needed to

[168] The study makes use of the *South Pacific and East Asian Corpus* (SPEAC; c. 1.52 million words; c. 200,000 words per variety), which comprises articles (news and editorials) from various newspapers available online. The data were collected during two periods in 2004 (for a detailed description of the corpus compilation process and various difficulties, such as the absence of information on authors or the potential inclusion of material from international press agencies, see Hundt & Biewer 2007: 251–252).

prove or disprove the hypothesis.[169] Further, they note that exonormative orientation still plays a part for L2 varieties of English, but that endonormative usage patterns (such as the use of the PrPf in SPst contexts) might emerge in these varieties, arguably modeled on the Antipodean L1 varieties (2007: 263). Finally, from a methodological point of view, their approach yielded some interesting results, but reliable automatic data retrieval and, more importantly, automatic identification of PrPf instances (e.g. using tagged corpus material) may help to avoid some of the difficulties and skewing effects that materialize in their study, which are due to the manual identification of the PrPf and the SPst in a smaller sample of the material and subsequent extrapolation to the full dataset (2007: 258–259).

While it has to be kept in mind that this study was predominantly a methodological exploration and thus the robustness of its findings should not be overestimated, its merits lie in the fact that it highlights, broadly speaking, the interesting insights that a study of the more "covert" features from the area of grammar can provide based on data from various varieties and, more specifically, the importance of variety type ("old" L1 vs. "new" L1 vs. L2) as a factor that should be considered in further studies on PrPf usage. Furthermore, it problematizes the validity of geographical proximity versus sociocultural influence (in the sense of the orientation of the L2 varieties; see also Hundt 2013) as factors determining variation. It has to be noted, however, that due to its restricted dataset (press texts), the study lacks information about effects of register or genre as well as about alternative surface forms (such as DONE-perfects, bare verb stems, etc.) that can express a PrPf notion.

3.5.5 van Rooy 2009

This investigation was based on one of the observations provided above, namely that the status of the PrPf has been widely debated for standard

[169] See e.g. Biewer (2008), who, also on the basis of data from online newspapers, investigates NZE as a potential epicenter for a number of Englishes from the South Pacific region and comes the conclusion that in the area of the PrPf, "NZE becomes a new model for the national standard in Samoa, Fiji and the Cook Islands" (2008: 215).

AmE and BrE but not for other varieties. In order to remedy this situation, van Rooy (2009) compares data from three components of ICE[170] (GB, EA, HK) with respect to the semantics of the PrPf construction (in the sense of the surface form HAVE + V-*en*). He tries to tackle the issue of categorization of the PrPf as either tense or aspect by determining whether there is a "shared core" (2009: 311), in other words, a prototypical meaning, across the varieties under investigation. The main rationale behind this approach is that taking into account findings from L2 varieties of English will lead to a deeper understanding of the grammatical features of English in general (2009: 312).

Although this work is principally of a corpus-based nature, it should be noted that van Rooy's findings are based on a comparatively small sample (600 occurrences) from two registers (dialogues and unpublished student writing), probably because they had to be extracted and coded mostly manually (2009: 313–314). The first part of the quantitative analysis finds the PrPf to be overall most frequent in BrE compared to the other varieties and most perfects to be PrPfs across all varieties (2009: 316–317). The ensuing semantic analysis is quite distinct from earlier approaches as presented in Section 3.3: it recognizes twelve different semantic readings, which are subsequently outlined and exemplified. The author claims that these twelve readings, of which the experiential reading is most frequent, can be grouped into two prototypical categories on a superordinate level, "continuative" and "non-continuative", with the latter being the most frequent and thus the most prototypical (2009: 317).[171] The remaining semantic readings (such as recent past, resultative, ongoing event as well as domain-specific usage, for example in academic and legal contexts, etc.) are interpreted as "elaborations" of the prototypes, while van Rooy argues that, among others, HAVE *got* constructions or the PrPf expressing a "past fact" constitute semantic categories of their own (2009: 319–328).[172] Another note-

[170] More detailed information on this corpus is available in Section 4.1 below.

[171] These findings echo the view expressed by Huddleston & Pullum (2002: 141) and corroborate the results in Wynne (2000) and Schlüter (2002a).

[172] Binnick uses a related, but slightly different approach, finding that existential and resultative are the core meanings of the PrPf, while continuative is less central and recent past is based on mere pragmatic inference (Binnick 1991: 99).

worthy finding is that the majority (c. 80%) of "prototypical" perfects occur without temporal specification by an adverbial (2009: 319),[173] which suggests (i) that temporal specification is an indication of a more marginal semantic meaning, and (ii) that the role of temporal adverbials might have been overestimated in earlier research. This finding also indicates that the PrPf in the varieties under investigation occurs with temporal adverbials to a lesser extent than in BrE or AmE, where only c. 65–70% of all instances are without specification (Elsness 1997; Schlüter 2002a; Hundt & Smith 2009). Due to the different methodological approaches and datasets used in these studies compared to van Rooy (2009), it is not clear whether this difference is of any significance.

The author concludes that, globally, the PrPf is used similarly in the three varieties under investigation, and states that some of the meanings of the PrPf are more central, usually conveying tense, while the more marginal meanings or "elaborations/extensions" usually express some kind of aspectuality. Therefore, he labels the discourse surrounding the grammatical status of the PrPf a "fake debate, which only exists in a context where a single or composite meaning [of the PrPf construction; V.W.] is assumed", and suggests that a theory that takes into account "shades of meaning" (2009: 329) from tense and aspect would be most adequate (see also Section 3.3.4).

While van Rooy's approach highlights the value of comparative studies involving both L1 and L2 varieties of English as well as the fundamental nature of the continuative versus non-continuative readings divide, for methodological reasons his findings have to be taken with a pinch of salt as regards the overall analysis of the PrPf. He only uses a very small sample, which leads to very low token counts for some of his semantic readings and renders the comparison between the varieties under investigation difficult.[174] In addition, the overall picture remains

[173] In tendency, this echoes Schlüter (2002a: 355) and other previous work, although the value for non-specified PrPfs is lower in these studies (60–70%).

[174] See for example the analysis of one type of domain-specific usage with fifteen occurrences for ICE-EA, three for ICE-HK and two for ICE-GB, based on which, van Rooy (2009: 322) infers (with caveats) a potentially "more entrenched" usage in EAfE.

restricted as his analysis considers data from three varieties[175] and two text types[176] only. At times the criteria underlying the identification of one type of meaning or another are not made explicit – for example, his semantic attribute "special associations with register or lexical items" (2009: 314) seems somewhat arbitrary – and may on occasion result in a conflation of variants of form and meaning, particularly with regard to the categories HAVE *got*[177] and progressive perfects. He also asserts that the difference between a prototype and one of the elaborated meanings is "a matter of degree" (2009: 320). This calls into question the value of these additional meanings, but at the same time emphasizes the validity of the continuative versus non-continuative distinction.

3.5.6 Hundt & Smith 2009

Hundt & Smith (2009) aim to put into perspective a number of previous claims and findings about the frequency of the PrPf in BrE and AmE. To this end, they trace its development in the second half of the 20th century

[175] Van Rooy notes that "[g]iven the nature of the data analysis, a comprehensive analysis of many different varieties is not feasible" (2009: 312). Although it has to be admitted that manually coding occurrences of the construction needs time and effort, the inclusion of at least one of the transplanted varieties for which comparable data are available (such as AusE, IrE or NZE) would have further increased the validity of the findings.

[176] The rationale for choosing dialogues and unpublished student writing is that these are supposedly the "least edited ones in the corpora, and therefore represent the local features of the language perhaps better than writing" (2009: 313). While this is doubtless true, written and more formal registers also provide further insights as they, too, are part of actual English usage in all varieties of English. Therefore, they should not be disregarded, especially in a study that ventures to identify core meanings of a particular construction. In addition, the potential influence of register differences cannot be accounted for; thus, the concluding statement that "[t]he core sense of the perfect clearly emerges across the various varieties and registers" (2009: 329) is something of an overstatement in light of the data on which the study is actually based.

[177] Although HAVE *got* constructions are among the most frequent in PrPf contexts (Biber et al. 1999: 463–467), the vast majority of surveys (the present study included) exclude them from their analyses as, arguably, they can always be replaced by a present form (Wynne 2000: 33; see also Hundt & Biewer 2007: 254) and thus no longer create reference to a past situation comparable to a true PrPf.

using the BROWN-family of corpora,[178] with a special focus on the relative frequency of the PrPf in contrast to the SPst and on co-occurring temporal adverbials.

They find that the overall frequency of the PrPf has decreased only slightly in the two varieties and, *pace* Elsness (2009b), that variation between them is stable, so that claims of a convergence of BrE and AmE are unfounded (Hundt & Smith 2009: 48). However, the decrease is more pronounced, i.e. statistically significant, in some text types (such as BrE press texts or AmE general prose) than others (Hundt & Smith 2009: 49–50). This suggests that genre differences may play a part in determining PrPf frequencies.[179] In passing, they also report that the relative distribution between the PrPf and the SPst has not changed (Hundt & Smith 2009: 51).[180] This comes as a surprise, as many previous authors (e.g. Elsness 1997) have taken for granted that the SPst has been expanding its territory at the cost of the PrPf (but cf. Wynne 2000: 45). With regard to co-occurring temporal adverbials, Hundt & Smith corroborate earlier studies in that the vast majority of PrPf constructions (approximately two thirds across all data) occur without temporal specification. Their findings also suggest that variation is relatively stable across varieties and time, but that the paths of development differ for individual adverbials (*already, never, ever, recently, yet*) in variable contexts (2009: 50–55). In addition, they note that it is not uncommon to find combinations of typical SPst adverbials such as *yesterday* or *last week* in combination with the PrPf in spoken data,[181] although occurrences in their written material are rare.

All in all, Hundt & Smith's results boil down to the conclusions (i) that variational patterns of the PrPf in PDE are due to regional and

[178] C. 1 million words each of written material from LOB (1960s) and F-LOB (1990s) for BrE; BROWN (1960s) and FROWN (1990s) for AmE. Cf. also Meyer (1995) for an earlier study on the PrPf using LOB and BROWN.

[179] A similar result is reported in a diachronic study of spoken BrE (Bowie et al. 2013).

[180] Cf. also Meyer (1995: 206), who provides slightly different numbers.

[181] They use parts of the BNC and the LCSAE for comparison and attribute these occurrences to two main sources: (i) an overriding pragmatic principle of current relevance *sensu* Rastall (1999: 81); or (ii) features of "online language production" such as unclear sentence boundaries or afterthoughts (Hundt & Smith 2009: 56; see also Werner 2013b and Section 6.6 below).

not diachronic differences (2009: 51), (ii) that variation between the PrPf and the SPst is an instance of "stable layering" (2009: 58), and (iii) that register differences may play a part (2009: 57). While their study again highlights the benefits of working with tagged versions of corpus material, it has to be pointed out that their findings are mainly restricted to written material from BrE and AmE and that the spoken corpora used for comparison are not perfect matches due to different compilation principles.

3.5.7 Davydova 2011

The goal of Davydova's (2011) monograph is to arrive at a comprehensive account of the PrPf in non-native Englishes. Her analysis comprises both learner and second-language varieties from a range of geographical locations and contrasts them with standard BrE.[182] As one of the main aims of her study is to identify cross-varietal sociolinguistic and language-internal factors that determine variation between the PrPf and other surface forms that express a perfect notion (in particular the SPst and, to a lesser extent, the Simple Present; see also Davydova et al. 2011: 300),[183] her analysis employs a semantic-pragmatic definition of "present perfect contexts" (Davydova 2011: 1).[184] That is, she analyzes contexts in which a PrPf construction expressing resultativeness, continuativeness, etc. would typically occur in standard English. She suggests that linguistic complexity is a theoretical backdrop for variation and may

[182] Davydova uses data from a large variety of sources, most notably informal spoken material from different components of ICE (EA, IND, SIN; category S1A; c. 200,000 words for IND, SIN; c. 90,000 words for EA), the *London-Lund Corpus of Spoken English* (LLC; c. 400,000 words of spoken BrE) and the *Hamburg Corpus of Non-Native Varieties of English* (HCNVE; between c. 14,000 and c. 37,000 words per variety), which is made up of sociolinguistic interviews. For illustrative purposes, data from the *Hamburg Corpus of Irish English* (HCIE; no word count given) are used. Relevant instances are manually identified and coded, and for the quantitative analysis, distributional and variable rule tables are provided; however, due to methodological constraints only the latter type is available for the PrPf-SPst paradigm (Davydova 2011: 140–142).

[183] In a similar fashion to Elsness (1997), she defines the PrPf as a tense, enabling her to categorize the SPst and the PrPf in the same way (Davydova 2011: 42–50).

[184] For a short discussion of approaches and potential pitfalls when defining a variable context see Van Herk (2010: 50–51).

hinder acquisition of the PrPf construction by learners.[185] She further notes that complexity varies in different non-native varieties in the sense that different numbers of surface variants may exist and that various factors may constrain the occurrence of one surface form or another in a variable context. In addition, she investigates the influence of second-language learning processes and the validity of current relevance in determining the choice between the SPst and the PrPf in both PrPf and definite past contexts (which typically require a SPst).

Due to the scope of the study, a wealth of quantitative and qualitative findings is provided; thus, only the key implications can be outlined here. From a quantitative point of view, a clear continuum emerges with regard to the variety of different surface forms that occur from standard BrE (as reference variety) to acrolectal, mesolectal and basilectal varieties (Davydova 2011: 292; esp. Figure 17.1), which suggests that surface forms other than the PrPf and the SPst do not occur in PrPf contexts in native(-like) varieties such as BrE and acrolectal IndE. In addition, the multivariate analysis of constraining factors shows that semantic context (resultative, existential, etc.) is an important factor in all varieties and the exclusive factor in the native(-like) varieties,[186] while Aktionsart exerts an influence in the mesolectal varieties. In general, the more factors involved in determining the choice between different surface forms, the more a certain variety is supposed to be distant from (standard) BrE (2011: 293–294). Another finding in line with previous research is that, in the majority of cases, the PrPf does not co-occur with temporal specification (2011: 304).[187]

[185] In other words, the more a variety succeeds in producing native-like PrPf usage, the more its speakers master a complex phenomenon and therefore the variety can be seen as more complex (Davydova 2011: 291). Cf. in that regard also Eckman's (1977: 320) concept of *typological markedness*.

[186] To a lesser degree, temporal specification by an adverb is an important factor in BrE (Davydova 2011: 294; 303–304).

[187] Note that Davydova conflates unspecified contexts with indefinite contexts, which are supposedly typical of the PrPf in contrast to contexts with definite temporal specifications that are characteristic of the SPst (Klein 1992). Therefore, her conclusion that indefinite past time reference is one of the central underlying notions of the PrPf as the majority of PrPf instances occur in temporally unspecified contexts (Davydova 2011: 304) is erroneous.

These findings lead to the claim that the PrPf is a linguistically complex phenomenon (see further Liszka 2003: 16; Payre-Ficout et al. 2009: 1),[188] that is acquired to different extents in different varieties. Davydova notes that patterns of PrPf usage are surprisingly similar across the geographically diverse varieties under investigation, which renders substrate influence an unlikely factor (see also Davydova et al. 2011: 305) and emphasizes the importance of underlying general learner strategies, such as avoidance strategies, in the face of a complex linguistic category (Davydova 2011: 301). Second, she highlights the importance of current relevance as a notion that has to be acquired in order to achieve native-like PrPf use. Another noteworthy finding is that, apparently, current relevance is a stronger semantic notion in resultative and continuative contexts than in existential and recent past contexts (2011: 302–305), which implies that the favored readings are more central manifestations of the PrPf (but cf. van Rooy 2009).[189] Yet, current relevance may be expressed by different surface forms in non-native varieties of English, for example by the present tense, BE-perfects or bare verb stems. In some cases, this may be due to the influence of the native language of the speakers under investigation (Davydova 2011: 308). Third, she suggests that L2 varieties of English share some features and processes with learner varieties, which implies that approaches from these two fields should be reconciled (2011: 308–309).

Davydova's (2011) study can be seen as pioneering as it is the first attempt to tackle the PrPf in non-native varieties of English from a broader perspective, and thus to identify common underlying factors that determine its variation. She also convincingly shows that the notion of complexity is a helpful instrument for explaining this variation. With

[188] It is evident that the non-native varieties are more complex in their range of surface forms that express a PrPf context and with regard to language-internal variation, as more constraining factors are involved, while acrolectal IndE is the most successful variety in terms of mastering the PrPf as a complex phenomenon in view of acquisitional difficulty (Davydova 2011: 295–297). This illustrates the observation that language contact may lead to complexification of a system (Thomason 2001: 64–65).

[189] Cf. the hierarchy of semantic readings of the PrPf that is established (Davydova 2011: 66). The graphical representation might be misleading as the triangle-shaped figure suggests that experiential and recent perfects have a wide range of usage, when in fact they are claimed to be less central manifestations of the PrPf.

regard to the second theoretical notion that is presented as crucial for the analysis of the PrPf, current relevance, she concedes that the dividing lines between the semantic (2011: 56) and pragmatic (2011: 305) parts of the concept are far from clear (2011: 74),[190] a fact that has been noted by many critics of the concept (see Section 3.3.1.2). By the same token, the inclusion of a formal component ("adverbials of the '+ current relevance' group"; 2011: 73) as a defining feature of the concept of current relevance seems impractical, particularly in light of the fact that temporal adverbials do not occur with the majority of PrPf instances (Wynne 2000; van Rooy 2009; Hundt & Smith 2009 amongst others; cf. Davydova 2011: 304). Even if this was not one of the main issues of Davydova's study, the choice of her polysemous[191] semantic model could have been accounted for in more detail (see also Section 3.3 above), particularly as alternative (corpus-based) approaches, such as those developed in Schlüter (2002a) and van Rooy (2009), are not considered.

From a methodological point of view, Davydova demonstrates the usefulness of smaller corpora and sociolinguistic interviews for the study of grammar, although some of the variants identified occur too rarely to be investigated quantitatively. The choice of her corpus material is restricted, although the s1a category in ICE is without doubt the best candidate in a search for vernacular features (a similar approach is taken e.g. by van Rooy 2009, Lange 2012 and Seoane & Suárez-Gómez 2013). On a related note, it is surprising that she uses BrE data from the LLC instead of the ICE-GB component, which arguably would have been

[190] It is symptomatic that current relevance in the course of the investigation is defined as both "pragmatic concept" (Davydova 2011: 305) and "semantic component" (2011: 63, 65). In support of the preservation of the fuzzy notion of current relevance, other authors maintain that it can be applied in empirical studies if tokens are coded for a number of relevant factors (such as adverbials) that can be interpreted as correlating with a PrPf surface verb form (Van Herk 2010: 51).

[191] Davydova wrongly claims that "polysemous" corresponds to "non-monadic" *sensu* Petersen (2004: 116). Rather than meaning that the PrPf has different semantic readings, "non-monadic" implies that features of the grammatical categories tense and aspect are united (see also Section 3.2.4). Note further that the claim that her polysemous approach (Davydova 2011: 55) is compatible with that of Declerck (2006) is somehow skewed, as the latter approach is also polysemous, but only on a lower level. The "core meaning or semantics" (Declerck 2006: 312) of the PrPf in his account is anteriority (as in McFadden & Alexiadou 2010: 392; see also Table 3.3.1).

more comparable. Furthermore, she does not disclose why the results from the three components of ICE (EA, IND, SIN) that are used for the quantitative analysis as instances of mesolectal varieties are not taken into consideration in the main overview of the findings (Davydova 2011: 291–305).[192] However, it is indisputable that Davydova (2011) has indicated the potential for extension of the data from other L1 varieties of English, for which ample corpus material is available,[193] to include additional factors that constrain PrPf contexts, such as preceding tense, and to consider register and text type differences, thus going beyond a mere description of spoken language.

3.6 Summary and discussion of implications for the present study

Sections 3.1 to 3.2 presented various views on the grammatical status of the PrPf (and other perfects) in English, starting with reference grammars and highlighting approaches such as "perfect as aspect", "perfect as tense", "perfect as a category of its own", and some alternative views. Based on the foregoing, it is clear that for the time being we must accept that the issue of the grammatical status of the PrPf is far from settled, although it is hoped that the previous sections may have helped to find one's way through (and systematize) the terminological jungle surrounding the PrPf in English.

---

[192] Davydova et al. (2011) provides a short analysis of (upper-)mesolectal varieties (data from ICE-EA, SIN, IND + further data from HCIE and mesolectal IndE). Parallel to the findings of Davydova (2011), this paper suggests that the semantics of PrPf contexts is the most influential factor that constrains variation and that, while all these varieties show similarities as far as variation between different PrPf surface forms is concerned, they predominantly use the PrPf construction. Therefore, Davydova et al. (2011: 304–305) conclude that for these varieties, substrate influence can be discounted and that long exposure to the target variety is one of the sociolinguistic factors that strongly determines orientation toward standard BrE.

[193] The same holds for AmE, which arguably exerts a strong influence on non-native varieties of English (e.g. via globalized popular culture) and may be responsible for the fact that the SPst is used in these varieties to a greater extent than in BrE. In more general terms, this would problematize the role of BrE as the "gold standard" for learners and second-language speakers of English. Yet other native varieties such as AusE might exert an influence on geographically or socioculturally close varieties (Hundt & Biewer 2007).

While a number of alternative approaches have been developed by different authors and the inherent weaknesses of both the "perfect as aspect" and the "perfect as tense" hypothesis have been acknowledged, notably the fact that both have to employ some kind of workaround to fit the PrPf into their general system of grammatical categories, they still seem to dominate the current discourse. In contrast, accounts of the "perfect as a category of its own" as well as further alternatives have not yet received the scholarly attention they deserve, although some of them (e.g. Kortmann 1991) provide straightforward and elegant alternatives to the two mainstream models.[194] For practical purposes, the PrPf hence is referred to as a form that creates temporal reference (TR form).

It has to be kept in mind, however, that the vast majority of the aforementioned theoretical considerations are based on the description of (standard) BrE and AmE,[195] whereas the grammatical status of the PrPf, or, more generally speaking, of surface forms with a perfect notion, may differ in diverse World Englishes compared to traditional native varieties. In addition, it has to be emphasized that it is not the aim of the present study to develop yet another theory on the grammatical status of the PrPf in English, as this is not the main focus of the empirical analysis below. The intention is, however, to arrive at some meaningful conclusions about its status in World Englishes and to explore whether the situation is similar or dissimilar across varieties or groups of varieties.

Section 3.3 provided an outline of previous studies and theories of the semantics and pragmatics of the PrPf, and discussed the different ways in which the PrPf can be viewed as a compositional construction. Similar to the issue of grammatical categorization, it was shown that a multitude of different accounts exist with respect to the semantic description of the PrPf. These claim either a single underlying meaning or a number of discrete meanings or readings. While monosemous accounts have been criticized for being too broad to allow any meaningful

[194] See also Bybee & Dahl's view from a typological angle that considers all attempts at categorization of the perfect as either tense or aspect as futile and suggests that "the relevant entity for the study of grammatical meaning is the individual gram" (1989: 97). The present analysis aims to proceed in this spirit.

[195] Phillipson (2009: 31) strongly criticizes this situation.

statements about the PrPf, polysemous accounts have the serious drawback that the individual interpretations are established on more or less subjective grounds,[196] as evidenced by the range of semantic functions (varying from two to seven, see Table 3.3.1 above) and the fact that cross-categorizations are often possible (Michaelis 1994: 123–126; Portner 2003: 459–460). This contradicts the view that there is a "general consensus" (Tagliamonte 2000: 338) in the literature with regard to the semantic categorization of the PrPf.[197] Other weaknesses of both the mono- and polysemous approaches include the fact that the proposed meaning(s) of the PrPf are established purely by way of introspection, which again reinforces the allegation that they are inherently subjective and only applicable to BrE and AmE, while different semantic functions may be salient in other varieties of English.

Pragmatic approaches provide further insight into the nature of the PrPf as they involve another level of linguistic analysis and recognize the importance of extralinguistic context, but, like the semantic approaches, they do not fully succeed in offering a unanimously accepted view. Notions such as current relevance can be retained provided they are elevated one level on the linguistic hierarchy (from semantics to pragmatics), but the fundamental problem of the lower level of that hierarchy (e.g. the arbitrariness in assigning different semantic readings) remains. Note also that "persistence", one of the proposed pragmatic values in Nishiyama & Koenig's (2010) model, corresponds to "continuative" in a semantic analysis and that, again, the establishment of the various pragmatic values is not immune to subjectivity. On these grounds the present study does not attempt to code the occurrences of the PrPf for pragmatic values.

As regards the various ways in which the PrPf has been described as compositional, it was shown that purely form-based approaches fall short for various reasons. Theories that view the PrPf as a combination

[196] But cf. Slobin (1994: 129), who suggests from the perspective of L1 acquisition that different readings of the PrPf may have an actual foundation in psychological reality, as the PrPf is used in different semantic contexts at different stages of the acquisition process.

[197] It is symptomatic that one of the more recent reference grammars of English (Biber et al. 1999) contains no attempt at a semantic classification of the PrPf at all, potentially to avoid the inconsistencies and disputes that surround this area of research.

of grammatical categories seem more adequate and intuitively plausible, but have to rely on additional contextual elements to explain its full range of usages. This implies that contextual elements (such as temporal adverbials) and other elements which are seen as partly constituting the meaning of the PrPf (such as Aktionsart of the verbs involved) are areas that are worth further investigation. The third type of compositionality, which sees the PrPf as a combination of a semantic and a pragmatic component, does explain co-occurrence patterns with temporal adverbials but, like the pragmatic approach criticized above, does not provide a convincing explanation overall for the characteristics of the PrPf.

In addition, Section 3.4 presented the findings from a number of studies on the PrPf in non-standard varieties. These findings complement the theoretical accounts in the sense that (i) they were able to identify further surface forms that can be used in contexts that require a PrPf in standard English, and (ii) they draw attention to the fact that the PrPf can be used in a wider range of contexts than in standard English. However, it was highlighted that the proportion of alternative forms in the non-standard varieties investigated is comparatively low, which suggests that the main focus of quantitative research should lie on occurrences of PrPf forms and their usage patterns. Furthermore, an extended functional range of the PrPf is attested for some varieties only, so it remains to be verified whether this development is idiosyncratic or whether it manifests itself as a stable trend in further varieties.

Section 3.5 discussed the results of some relevant corpus-based studies on the PrPf that mainly addressed cross-variational issues. While the strengths and weaknesses of the individual works were outlined, it emerged that (apart from Davydova 2011) the focus of book-length treatises lies on the BrE versus AmE paradigm only and that studies of the PrPf in further varieties of English are relatively restricted in their scope, being based on comparatively small datasets and/or material from a limited number of different genres only, usually excluding written texts. It also emerged that many empirical studies simply rely on traditional semantic categorizations that are based on introspection and may not be fully adequate. In addition, the occasional use of unbalanced corpus material leads to odd results in comparison to the findings of other studies, as does the use of untagged corpus material when searching for the

PrPf. This in turn highlights the importance of using comparable corpus material that is reliably tagged.

A number of additional points relevant to the present study are worth noting. First of all, it is obvious that grammar is a particularly interesting area of study, including varieties of English other than BrE and AmE. Second, it remains to be tested whether clear dividing lines between L1 and L2 varieties can be established and whether geographical and/or sociocultural distance plays a part in determining the similarity of the varieties in question. Third, some interesting hypotheses with regard to the role of particular linguistic categories and items such as Aktionsart, semantic context, and co-occurring temporal adverbials were posed, which either can be corroborated or rejected based on the observations of the present study. Fourth, the need for a contrastive analysis of differing registers became apparent. It should be evident that this is only possible with comparable corpus material across different varieties as employed in the present study. Fifth, apart from the pilot study by Hundt & Biewer (2007), none of the PrPf studies discussed above included data from transplanted L1 varieties such as IrE, AusE, CanE, or NZE, meaning that a potential orientation of L2 varieties toward one or more of these has not been considered in detail so far.

From a theoretical point of view, the corpus-based works only partly contribute to resolving the salient issues with regard to the PrPf. Although most studies discuss the grammatical status of the PrPf as a tense or aspect, alternative approaches such as Kortmann's (1991) suggestion are rarely considered, and theory is interpreted in such a way as to suit the respective author's research needs; for example, Elsness (1997) and Davydova (2011) (legitimately) define the PrPf as a tense, as their research focus is on the alternation with the SPst (whose status as a tense is beyond doubt). While the grammatical categorization is only of secondary importance for the present study, it should be noted that also with regard to the semantic analysis, supposedly canonical models are occasionally employed despite their inherent weaknesses, and studies rely on the by no means undisputed notions of indefiniteness and current relevance as characteristics of the PrPf (see sections 3.3.2.1 and 3.3.1.2). Notable exceptions are Schlüter (2002a), van Rooy (2009) and, with caveats, Wynne (2000), who refrain from categorizing the individu-

al occurrences of the PrPf in their data according to established classifications but rather develop models of semantic categorization based on their datasets. These corpus-based models correspond nicely in that they all identify "continuatives" versus "non-continuatives" as fundamental readings of the PrPf (see also Huddleston & Pullum 2002: 141–146), which implies that this distinction actually mirrors the central semantic categories that are discernible in the PrPf, and that applying them to other corpus material would be fruitful.

4 Methodology and data

The present chapter represents the first part of the empirical section. It provides an overview of the methodology and the data. In passing, it highlights problems and difficulties that occurred during the process of data analysis as well as advantages and improvements in comparison with other studies. Specifically, after a few notes on the layout of the corpus material, which serve to make the particular suitability of ICE for the purposes of the study clear, it explicates the repercussions of the processes of tagging the data and describes the identification and sampling of instances of the PrPf. In addition, the present chapter considers the coding of the factors that are analyzed, the implications of the statistical techniques that establish a measure of similarity across the varieties under investigation, and potential caveats.

4.1 The ICE project

In order to understand why the ICE family of corpora was chosen for the current study, this section provides a brief description of the ICE project and of the design principles followed in compiling the components of the corpus.

The idea for establishing a comparable family of corpora for a vast range of national or regional varieties of English first emerged at the end of the 1980s and is outlined in detail in a volume edited by Greenbaum (1996a). At present, twelve components are publicly available[198] and twelve more are under compilation (www.ice-corpora.net/ice/index.htm). The rationale behind the project is that meaningful compara-

[198] These are the components for Canada, East Africa (Kenya and Tanzania), Great Britain, Honk Kong, India, Ireland, Jamaica, New Zealand, the Philippines and Singapore. For the main part of the present study, the East Africa component (ICE-EA) was excluded as its textual categories for various reasons do not correspond to the general layout, which precludes an analysis of text type and register effects. Similarly, ICE Sri Lanka (ICE-SL) and the American component (ICE-USA), for which to date only the written parts are available, are not considered. Instead, it was possible to include data from the Australian component (ICE-AUS), as kindly provided by Adam Smith (Macquarie University, Sydney).

tive analyses of the varieties included are only possible with parallel corpora that all adhere to the same principles of design and thus are as homogeneous as possible with regard to the specification of text categories (see below), the dating of the data (mostly early 1990s) and the educational background of the speakers involved[199] (for more details see Greenbaum 1996b: 5–6). Each of the components comprises approximately one million words that are composed of 500 (ideally self-contained) texts of 2,000 words each, while many of the 2,000-word text units are composite themselves. The specific text categories are presented in Appendix A (Nelson 1996: 28–33 and www.ice-corpora.net/ice/design.htm).

Although by contemporary standards one million word corpora seem to be small, the main advantage of the ICE corpora is their comparability across an extended number of varieties. From a methodological point of view, this enables the researcher to undertake (i) investigations of one particular variety, (ii) comparative studies with a focus on the proximity to or distance from a reference variety, most likely BrE or AmE, (iii) analyses of potential "core" features across varieties and (iv) studies of variety-internal variation. Recent studies, echoing Greenbaum, have emphasized that it is crucial in cross-variety analyses to have datasets that are comparable as far as possible. This is especially important if variation across genres or text types should properly be mirrored with-

[199] Note that all informants are adults with at least a completed formal English-medium secondary-school education. They are considered speakers of "'educated' or 'standard' English" (Greenbaum 1996b: 6) of the respective variety. Occasionally speakers are included due to their social or public status (e.g. writers, broadcasters, etc.). Therefore, the data in the corpus material is relatively controlled for social variation, which is usually comparatively more pronounced in typical L2 varieties (Schneider 1983a: 73–74 for JamE). However, the compilers should aim at representativeness as to sex, age and regional descent of the speakers (Nelson 1996: 28). Note further that, as more and more components of ICE have been added at a later stage (e.g. the compilation of the Ugandan component was started in 2011), the dating criterion had to be relaxed, necessitating a note of caution: For a comparison of older ICE components (such as ICE-GB or ICE-NZ, data from the early 1990s) with ones to be completed in the future, say, in 2020, researchers have to be aware of potential effects of diachronic variation (cf. studies based on the BROWN-family of corpora, also addressing variation over a 30-year span, data from the 1960s and 1990s). A possible solution would be to compile new ICE components, facilitating diachronic analyses (see also below).

in that analysis (Hundt & Vogel 2011: 146). In the same vein, Mukherjee & Schilk (2012: 194) have pointed out that conceiving of worldwide varieties of English as monolithic blocks may be too reductive a view and, further, that taking account of variety-internal variation is a growing and necessary trend in research of World Englishes. Still, caution is needed due to the various processes (language contact, koinéization, etc.) that are involved in shaping varieties of English (see Chapter 2; cf. Davydova et al. 2011: 293).

Another argument for the applicability of seemingly small corpora such as those from the ICE family can be found in the nature of the current study as an analysis of grammatical patterns. Hundt & Mair (1999: 224; see also Biber 1990) argue that, in contrast to lexical studies that rely on extensive monitor corpora (like, for example, COCA; Davies 2010) to include as many data and in consequence words as possible, one million word corpora provide enough evidence for grammatical studies. This finding is still considered valid and holds especially for items with a comparatively high frequency of occurrence such as the PrPf.

A further aspect worth mentioning pertains to the establishment of textual categories. Not only can they be discerned according to the mode of discourse/channel of production (written vs. spoken; cf. above), but also along other dimensions. In the present study I will also use a distinction between "involved" and "informational" texts, first established in Biber (1988) as "Dimension 1" along which texts may vary. Applied for the ICE genres, dialogues, monologues and non-printed writing are subsumed under the former, printed writing under the latter category.

It has to be acknowledged, though, that ICE categories do not match Biber's categories one-to-one, so some of the text types mentioned are part of a grey area between the two poles. For instance, in Biber's analysis (cf. 1988: 128) professional letters appear slightly below the zero value in the relevant figure, that is, toward the informational pole. In ICE, however, professional letters form a category (w1b) with social letters, which are high on the involved score. It could be argued that both subtypes are characterized by typical features of involved texts, such as a high incidence of first and second person pronouns (see further Biber &

Conrad 2009: 247). Depending on the topic, student writing (w1a), which does not form part of Biber's analysis, can display features classified as involved, so this is why non-printed writing is also seen as located toward the involved pole of the dimension. In addition, it could be argued that the ICE categories are internally inhomogeneous as, for example, scripted monologues (s2b) contains TV and radio news, where linguistic features may differ and which also consist of different phases (exposition and argumentation, which is more involved, and narration, which is more informational). In any case, the empirical analysis will show that the application of Biber's categories is helpful.

4.2 Tagging, identification of Present Perfect occurrences and sampling

The review of earlier studies on the PrPf (see Section 3.5) revealed that reliable identification of PrPf occurrences is only possible when using corpus material that is tagged for parts of speech. Otherwise, a disproportional amount of manual work is needed and, if conclusions are based on extrapolations from a small sample that was manually retrieved, the results may be skewed. As at the data analysis stage of the present study ICE-GB was the only component for which a tagged version was available (only analyzable with ICECUP; www.ucl.ac.uk/english-usage/resources/icecup/),[200] the first step toward obtaining reliable data was to create tagged versions of the corpus files, which are usually available in plain text format.[201] This was automatically done with the help of the C5 tagset (ucrel.lancs.ac.uk/claws5tags.html) of the CLAWS part-of-speech tagger (ucrel.lancs.ac.uk/claws/). As regards the reliability of this tagger, the low 1.5% error rate that was established for the tagging of the BNC (see www.natcorp.ox.ac.uk/docs/URG/posguide.html)

[200] Versions of the ICE-components that are tagged with the C7 tagset as well as with semantic tagging are available now (www.ice-corpora.net/ice/index.htm).
[201] For the sake of consistency, a plain text version of ICE-GB was created that was subsequently tagged again.

could be confirmed for the ICE data with the help of spot checks.[202] For the identification of occurrences of the PrPf in particular, tagging errors[203] such as (1), where the form of the cliticized third person singular forms of BE and HAVE is congruent and wrongly analyzed by the tagger, (2), where the *-ing* suffix was interpreted as adjectival, (3), where an adjective was interpreted as verbal form[204] or (4), an instance of homography between a verbal form and an adjective could have skewing effects and had to be manually removed as false positives (after sampling). However, the rare instances of examples with non-standard concord between auxiliary and subject, such as (5), were included.

(1) whatever_DTQ it_PNP1 's_VHZ (correct: _VBZ) called_VVN (ICE-AUS s1b-019)
(2) authorities_NN1 have_VHB been_VBN pussy-footing_AJ0 (correct: _VBG) around_AVP (ICE-NZ w2e-009)
(3) he_PNP has_VHZ mixed_VVN (correct: _AJ0) feelings_NN2 about_PRP his_DPS daughter_NN1 's_POS marriage_NN1 (ICE-PHI s2b-006)
(4) You_PNP have_VHB lead_VVN (correct: _AJ0) ones_NN2 as_AV021 well_AV022 (ICE-JA s2a-054)
(5) Ah_ITJ you_PNP has_VHZ been_VBN to_PRP Vietnam_NP0 (ICE-HK s1a-007)

The second step after tagging was to retrieve all instances of the PrPf in the corpora. Before this could be done, a number of decisions as to which occurrences of PrPfs to in- or exclude from the study had to be

[202] See also van Rooy & Schäfer (2002), who found that the accuracy of the CLAWS tagger exceeded 96% for unedited and 98% for edited learner data where spelling errors were manually removed.
[203] Inaccurate tagging might sometimes also be due to idiosyncrasies in the transcription of spoken material in the corpora, e.g. I_PNP wouldnt_VVB (correct: would_VM0 n't_XX0) have_VHB done_VDN (ICE-JA s1a-028).
[204] Ambiguous cases such as *you have isolated exceptions to the rule* (ICE-PHI s1a-071) or *this new mode of peripheral specialization has limited developmental potential for Jamaica* (ICE-JA w2a-036) were also excluded, as were occurrences where the PrPfs formed part of a "false start" and were subject to self-corrections such as *kidnapping has been has been uh is already syndicated* (ICE-PHI s2a-046), which did not lend themselves to an unequivocal analysis.

taken. First of all, it was decided to exclude the highly frequent combination HAVE *got* (+ NP), as it can virtually always be replaced by a present form HAVE + NP (Wynne 2000: 33; Hundt & Biewer 2007: 254); likewise, all instances of HAVE *got to* as a semi-modal that expresses obligation (Ota 1963: 51–53) were excluded. In addition, to sharpen the focus of the analysis and to ensure comparability with other works, progressive forms were, with the exception of the case studies, excluded due to their almost exclusive association with continuative contexts (see also Meyer 1995: 222), although they constitute an interesting area and should be investigated in detail in another study.

The search then was done with the help of the concordancing software *WordSmith Tools* (version 5.0.0.334; Scott 2011). A search routine that involved the tags from the CLAWS tagset was devised[205] and each individual corpus component was searched, leading to a file with concordance lines including a substantial amount of context (c. 200 characters) before and after the individual occurrences of the PrPf for each component. After the manual removal of passive forms, which could not be identified by the automated search at this stage, representative random samples were drawn (98% confidence level, 5% margin of error)[206] and sampled lists of concordance lines were created and saved. These lists were manually scanned for false hits (mostly due to tagging

[205] The search was for the tags *_VHB/*_VHZ + *_VBN/*_VDN/*_VHN/*_VVN within four words to the right. The four-word range was deemed adequate as earlier studies (Schlüter 2002a: 103; 2006: 136) found that more than three inserted items between auxiliary and past participle are very rare. The following tags were excluded in the automated search according to the exclusion criteria: *_VBG/*_VDG/*_VHG/*_VVG/got_*. It should be evident that, unlike in other studies (e.g. Davydova 2011) the construction under investigation in this search is defined in purely formal terms. It has to be kept in mind, however, that there may not be a discrepancy in form but in function in some varieties of English in contrast to others (particularly in L2 varieties; Hartford 1993: 1), so caution is needed when looking at the individual occurrences (see also Section 6.4).

[206] The required sample size for each ICE component was calculated with R (version 2.14.1; see cran.r-project.org for the latest version) with the help of the *sample.size* function from the library *samplesize* (news.mrdwab.com/2010/09/10/sample-size-calculator-for-r/). Note that it was decided to keep the actual sample bigger than required in order to compensate for potential unwanted data (e.g. due to tagging and searching errors) that had to be removed manually afterwards. This proved beneficial as the *WordSmith Tools* search indeed returned a number of unwanted data points.

errors; see the examples above) and further unwanted data such as combinations of modals and PrPfs (Winford 1993: 172–174), elliptical forms and non-finite forms of the type *she is believed to have left,*[207] or occurrences with passive progressive forms that had not been excluded automatically by the tagger due to intervening material such as (6) or (7), which were all removed.

(6) The council will no longer be dominated by appointees, who *have been for years toeing* official lines (ICE-HK w2a-011)
(7) Again Christianity *has been really uh using* various ways (ICE-JA s2a-031)

These adjusted lists were subsequently exported to spreadsheets in order to proceed with the identification and annotation of relevant instances and the manual coding of each individual occurrence according to the factors outlined in the following sections. All in all, this resulted in the following breakdown of the 5,334 data points:

Table 4.2.1. Sample sizes and invalid occurrences in the ICE data

corpus name	PrPf occurrences	recommended/ actual sample size	invalid occurrences in the sample		sample size available after adjusting
ICE-AUS	4,167	480/523	7	(1.3%)	516
ICE-CAN	3,028	460/542	14	(2.6%)	528
ICE-GB	3,719	473/563	26	(4.6%)	517
ICE-HK	4,040	478/570	20	(3.5%)	550
ICE-IND	4,585	485/580	2	(0.3%)	578
ICE-IRL	3,188	463/549	19	(3.4%)	530
ICE-JA	2,828	455/555	13	(2.3%)	542
ICE-NZ	4,530	484/579	10	(1.7%)	569
ICE-PHI	2,746	453/535	8	(1.5%)	527
ICE-SIN	1,784	416/484	7	(1.4%)	477

[207] See e.g. Bowie & Aarts (2012) for an in-depth study of infinitival and modal perfects in BrE. Note that if PrPfs are embedded under a modal (Schaden 2009: 118) or an infinite construction (McCawley 1971: 100; cited in Mugler 1988: 62–63) the definite temporal adverbial constraint is overridden (see also Section 6.5 below). Constructions with modals were therefore excluded from the results of the main part of the empirical analysis.

4.3 Coding of factors

Once the sample that forms the basis for the current investigation had been created, the individual occurrences of the PrPf had to be manually coded for a number of factors that are either found in the context of the PrPf (temporal adverbials, sentence type, preceding tense), that are inherent in elements of the PrPf construction itself (Aktionsart of the main verb) or that are found on the level of semantics.[208] These factors were chosen as they had been identified as determining variables for the PrPf in earlier work (see e.g. Sharma 2001: 353 or Tagliamonte 2000: 340–347; cf. the detailed list in Van Herk 2008: 57 and Section 3.6) and thus seemed apt for establishing a kind of grammar of usage (cf. also Biber & Conrad 2009: 216) for the PrPf. All in all, 31,896 manual coding decisions had to be made.

4.3.1 Temporal adverbials

In spite of the finding that temporal adverbials overall only co-occur with roughly one third of PrPf occurrences,[209] they have been widely recognized as important triggers for the PrPf (in contrast to the SPst) in English since at least Jespersen (1931: 61).

It has been stated that, cross-linguistically, (i) temporal adverbials create reference to a point or stretch on the time axis (Klein 2010: 1224; and, for the English PrPf in particular, this reference must be to the present in one way or another; cf. Dinsmore 1991: 101; Leech 2004: 44–45), (ii) they are used less systematically to express location in time in

[208] This eclectic approach seeks to reconcile the two positions that either include or exclude contextual factors in their description of the PrPf, therefore aiming for comprehensiveness as far as factors from different levels of analysis are concerned.

[209] Cf. also Portner, who maintains that "any sentence which lacks overt temporal adverbials receives a default temporal specification, something like *once* or *at some time*" (2003: 492). While this view stands in stark contrast to earlier positions that saw reference to the present as the distinguishing criterion of the PrPf in opposition to the SPst (e.g. Koziol 1958: 499) and contains some inherent vagueness, it provides a rationale for the claim that the PrPf is usually used in indefinite past contexts. It remains to be tested whether this "default" reading of PrPf occurrences without temporal specification applies for all of the varieties under investigation.

comparison to other means such as tense or aspect (Dahl & Velupillai 2011a); in other words, they are less grammaticalized, and (iii) they can be seen as a wholly separate system for expressing temporality (Panitz 1998: 275). As regards the last point, however, it should be evident that reciprocal effects between the different means of expressing temporality exist (Binnick 1991: 300). Temporal adverbials rather serve to specify the orientation in time of an utterance in more detail than would be possible with tense or aspect alone (Schopf 1984: 403; Miller 2004a: 235; Jaszczolt 2009: 91), which renders them "by far the most elaborate device to encode time" (Klein 2009b: 64).[210] In addition, it has to be noted that, at times, the usage of a particular temporal adverbial can even determine other factors, such as the semantics of a PrPf construction (Guenthner 1977: 95; Vermant 1983: 43; Kirsten 1994: 54–55; Michaelis 1994: 132–134; Schlüter 2002a: 203), for example when a durational adverbial strongly favors a continuative semantic reading.

The actual coding of the data with regard to temporal adverbials was done in two steps. First, each PrPf occurrence was coded for presence or absence of an associated temporal adverbial (that is, within the scope of the respective utterance/verbal element). This makes it possible to establish the proportion of PrPf occurrences with and without temporal specification and to compare the values for the current dataset with findings from earlier studies. If the context that was extracted around each PrPf occurrence to the spreadsheet did not allow a straightforward analysis (e.g. due to an extended sentence length), the concordance files were consulted again to look at additional contextual data. Second, if temporal specification was present, the type of specification was determined. I followed the schemes proposed by Quirk et al. (1985: 8.55–8.73; 8.97) and Biber et al. (1999: 777), which recognize four groups of temporal adverbials according to their meaning.[211] These are:

[210] It has been noted for the Past Perfect that "absence of adverbial marking may facilitate the reinterpretation of pragmatic meanings of aspectual constructions" (Sharma 2001: 358). Again, this emphasizes the role temporal adverbials play in specifying the location in time of an utterance.

[211] Other schemes for categorizing temporal adverbials have been proposed for instance by Schopf (1984: 16); Binnick (1991: 306–309); Panzner (1995: 37); Leech (2004: 45–46) or Declerck (2006: 132–146, 591–633).

(i) Adverbials of time-position (TP; e.g. *today*, *yesterday*, *afterwards*, etc.)
(ii) Adverbials of span and duration (SP; e.g. *briefly*, *since* X, *for* X, etc.)
(iii) Adverbials of frequency (FR; e.g. *daily*, *twice*, *always*, *often*, *never*, etc.) [212]
(iv) Other adverbials of sequence or time relationship between two events (SQ; e.g. *already*, *originally*, etc.) [213]

It has repeatedly been stated (e.g. by Binnick 1991: 300–310; Panitz 1998: 247–271 or Klein 2009b: 64) that these four groups can be realized by a plethora of surface forms[214] also comprising compound constructions such as prepositional phrases, as in (8), subordinate clauses, as in (9), or noun phrases, as in (10), while single adverbs represent the most frequent type of realization (Schlüter 2002b: 310).

(8) The plunder that has gone on *in this period in which the trade unions have been weakened* [...] (ICE-GB s2b-036)
(9) *Since we last communicated*, SCI-MED has employed a service engineer in Wellington, so the after-sales support we offer is now better than ever before. (ICE-NZ w1b-022)
(10) Uhm *this year* I have taken six courses (ICE-HK s1a-058)

With the help of the information from the coded data, it can be determined whether varieties of different types pattern differently in their usage of temporal adverbials and whether significant differences between L1 and L2 varieties of English exist. Another hypothesis with ref-

[212] An alternative label for this group is "pattern adverbials" (Vlach 1993: 251).
[213] Occasionally, this group is also labeled "adverbials of contrast" (Klein 2009b: 65–66). Note that this category is absent from earlier work by the same author (Klein 1992: 528) and is also sometimes ignored in other works (e.g. Rothstein 2007: 97).
[214] For other lists of adverbials emerging from corpus studies, categorizations of these on formal principles and related discussion see e.g. Vermant (1983: 72–84); Panzner (1995: 107–134); Meyer (1995: 220–222); Wynne (2000: 52–113); Schlüter (2002a: 225; 2002b: 310–312) or Elsness (2009a: 98).

erence to the PrPf and usage of adverbials is the alleged predominance of occurrences that are not temporally specified in written and more formal genres (Miller 2004a: 231).[215] If confirmed, this would imply a higher rate of temporal specification in the spoken data. *Pace* Miller (2004a), it has been claimed, however, that patterns of temporal specification are relatively homogeneous both across various genres and across varieties (BrE and AmE), while they interact with sentence type (55% of all negative statements but only 25% of all questions are specified by a temporal adverb) and semantics (where overall less frequent semantic readings, such as continuatives, are more likely to be temporally specified) (Schlüter 2002a: 259–260).

In the present study as well as in earlier publications (Bauer 1970; Dahl 1985; Mugler 1988; Kortmann 1995; Portner 2003; Sempere-Martinez 2008; Ritz & Engel 2008; Elsness 2009b or Schaden 2009, to name but a few), it has often been stated that certain restrictions apply with regard to the combinability of the English PrPf with members of the group of temporal adverbials that is labeled (i) above, notably with those that express definite past time reference (such as *a year ago, yesterday, in 1974*, etc.), and various syntactic and pragmatic explanations for this finding have been discussed (e.g. by Klein 1992; Giorgi & Pianesi 1997; Portner 2003 or Rothstein 2007; 2011). However, when looking at actual language data, this view oversimplifies matters as these constraints do not always seem to be operative. Above all, AmE is commonly regarded as more flexible in terms of the choice between the PrPf and the SPst when co-occurring with temporal adverbials (Quirk et al. 1985: 4.22); i.e. in this variety the relative frequency of the SPst in contrast to the PrPf is higher in instances that involve an indefinite temporal adverbial, particularly in non-continuative contexts (Meyer 1995: 224) or at least with a restricted group of adverbials such as *already* or *just* (Panzner 1995: 146; Hundt & Biewer 2007: 253; see further Section 6.5).

[215] However, a special situation seems to hold for narrative texts. It has been stated that this genre, which usually implies fictionality and more often than not pastness, needs more temporal specification beyond tense and aspect than other texts, where the exact temporal reference may be implicitly clear for speaker and listener (Weinrich 2001: 91). On a more general level, this suggests that different text types potentially reveal different patterns of distribution of temporal adverbials (see furthermore Panzner 1995: 147).

In addition, a number of corpus studies of BrE and AmE have identified instances of the PrPf with co-occurring definite temporal adverbials in both spoken and written material (Ota 1963: 46; Meyer 1995: 225; Hundt & Smith 2009: 55–57), which cannot all be simply attributed to performance errors or inadvertence on the part of the speaker (Quirk et al. 1985: 4.23; Harder 1997: 417). It has been suggested elsewhere (Rastall 1999: 81–83; Davydova 2011: 156–158) that the principle of current relevance may supersede the combinational constraint for pragmatic reasons.[216] As regards their overall frequency, examples of this kind are comparatively rare, but seem to be more salient and systematically used in other varieties, such as NZE (Quinn 1999: 196; but cf. Hundt 1998: 74) or AusE (Ritz 2010: 3403; see further Ritz & Engel 2008), and further in varieties of English in the Asian sphere (Kachru & Smith 2008: 91–92; Balasubramian 2009: 92–93).[217] Thus, it remains to be tested whether and to what extent combinations of the PrPf and adverbials of the "definite past" group surface in the current dataset and whether there are differences between individual variety types and text types (see in particular Section 6.6 and Werner 2013b). Besides these combinations, arguably due to substrate influence, a number of (particularly Asian) L2 varieties merely rely on temporal adverbials for coding utterances for tense and aspect or dispense with coding reference to the past by morphosyntactic means (i.e. use the base form of verbs) after past time reference has been established at least once (Platt et al. 1984: 69; Kachru & Smith 2008: 91–92).

4.3.2 Aktionsart

The second factor found to be influential for the usage and interpretation of the PrPf and for tenses and temporal information in general (Bennett 1978: 33; Michaelis 2006: 231; Jaszczolt 2009: 86) is Aktionsart or inherent lexical aspect of the main verb involved.[218] The rationale

[216] For similar explanations relating to pragmatic pressure as a reason for the overriding of the combinational constraint see Miller (2004a: 235) or Hundt & Smith (2009: 58).
[217] For the distribution of the feature in other varieties see also eWAVE Feature 100 (2012).
[218] For a working definition of Aktionsart see Section 2.1 above. It should be evident that Aktionsart depends on the lexical meaning of a verb, which moreover determines "whether

behind this is that not only contextual factors are important. For instance, the majority of PrPf occurrences remain without (external) temporal specification and, therefore, the meaning "is in the majority of cases inherent in the verb phrase itself" (Schlüter 2006: 143; see further Winford 1993: 175). In addition, Aktionsart arguably helps to carry pragmatic information, such as signaling to the hearer that a situation has finished or is still valid at the point of speech (Bauer 1970: 190; Ritz & Engel 2008: 135). However, this may also be attributed to the lexical meaning of the verb itself, which nicely illustrates the location of Aktionsart somewhere between grammar and lexicon (Dahl & Velupillai 2011a). At the same time, it has been stated that conditions of PrPf usage in general are "independent of individual semantic properties of each lexical verb" (Pietsch 2009: 537). The validity of the latter point is to be discussed in light of the data obtained as others have argued that the semantics of the PrPf may in fact in some cases be determined by Aktionsart of the verb involved (Naumann 1998: 193; Allerton 2008: 27). For example, continuatives can be determined by stative Aktionsart, as in (11).[219]

(11) Babylon's prices have remained reasonable (ICE-JA w2b-003)

For the coding of the data for Aktionsart, a large number of potential schemes are available. While some only employ a simple binary distinction between static and dynamic verbs (Bybee et al. 1994: 55), many – e.g. the elaborated scheme with 19 different predicate types presented in Meyer (1992: 183) – are quite intricate, and have been criticized as being "very specific" and "unsystematic" (Binnick 1991: 170) or as causing classificatory problems as they are too fine-grained and thus potentially create conflicts by allowing too many cross-categorizations (Klein 1994: 80). A relatively simple but commonly accepted and widely-used scheme

---

or not the action has led to tangible results" (Bauer 1970: 190). It is acknowledged that the boundaries of the individual Aktionsart categories as recognized in the present study may be fuzzy at times (see also Klein 2010: 1232).

[219] See also Portner (2003: 493). In a similar fashion, Ota does not relate to Aktionsart directly but views semantics as mere "reflection of the lexical meanings of the verbs that go along with the perfect form" (1963: 57–58).

(e.g. by Schopf 1984; Binnick 1991 or Klein 2009b) is that developed by Vendler (1957; 1967: 102–103), and this is also the one applied to the current data. It distinguishes four basic types of Aktionsart:[220]

(i) Activity verbs (e.g. *run, eat, fly*, etc.)
(ii) Accomplishment verbs (e.g. *build, draw* (*a circle*), *run* (*a mile*), etc.)
(iii) Achievement verbs (e.g. *discover, reach* (*the top*), *cross* (*the river*), etc.)
(iv) State verbs (e.g. *love, hate, be,* etc.)

The four categories are determined by the features stative versus non-stative, telic versus atelic (temporal boundedness, i.e. they possess a beginning and/or an end), short versus long duration, and potentially inner quantification and "phase" (inchoative/ingressive, terminative, continuative) *sensu* Klein (1994: 79–80; 2009b: 59–64).[221] Based on these features, the following rules helped to categorize the examples: (i) state verbs and achievement verbs usually would not allow a progressive form, while progressives are in principle possible for activities and accomplishments; (ii) state verbs are usually durative whereas achievement verbs are instantaneous;[222] (iii) activities are usually atelic whereas accomplishments are telic; (iv) accomplishment verbs usually imply some kind of build-up while achievement verbs are punctual.

From the examples given above it is clear that Aktionsart does not solely depend on the verb itself but is occasionally sensitive to context,

[220] Comrie (1976) added the category of "semelfactive" (exemplified by punctual event verbs such as *knock, sneeze* or *blink*). For the present study, it was decided not to include this category as in the full sample of PrPf occurrences only one instance occurred that would qualify as potential semelfactive verb (*I don't know how long the phone hasn't ringed*; ICE-HK w2f-011).

[221] The tests for Aktionsart as provided in Katz (2003) proved helpful in assigning a category in doubtful cases. See further Croft (2012: 33–45) for a critical review of Vendler's classification.

[222] As a matter of fact, this criterion for distinction might be seen as precarious: as rightly noted by Klein, "in reality there are no situations without duration" (Klein 1994: 33). A lowest common denominator could be that state verbs usually have a considerably longer duration than achievement verbs and this should suffice to categorize them contrastively.

for example an accompanying object realized by a noun phrase or modification by an adverbial. This is nicely illustrated by the following examples adapted from Smith (1995: 105–106), who argues that a differentiation between a "basic-level" and "derived level" categorization is necessary:

(12a) He swam in the pool.
(12b) He swam 100 meters.
(12c) He swam to the small island.
(13a) Kim knew the truth.
(13b) Suddenly Kim knew the truth.

While *swam* in (12a) can be interpreted as activity on the basic level, the context suggests an accomplishment interpretation in (12b) and (12c). A similar situation holds for the second example *knew*, which in (13a) represents a typical state verb, while in (13b) it can be interpreted as an achievement on the derived level due to the adverbial modification.

For reasons of comparability with other studies it was decided to adhere to a derived-level approach that takes account of contextual modifications and great care had to be taken during coding to identify the applicable Aktionsart for each individual PrPf occurrence. The benefit of this approach is that it provides a more accurate representation of the inherent semantics of the individual tokens and thus mirrors actual usage more closely. The coding for Aktionsart makes it possible to establish the distribution of the individual Aktionsart categories in the components and moreover to establish interplay between Aktionsart and other factors.[223]

[223] The general distribution and frequency of verbs may also play a role. Large-scale corpus studies (Biber et al. 1999: 464–465) have found that particular verbs are much more likely to occur with the PrPf than others, but that, as a rule, these are verbs that are highly frequent in general. Differences between registers exist, however. For instance, activity verbs are the most common and state verbs the most uncommon verbs with the PrPf in general; but the distribution between verbs of different Aktionsart categories was revealed to be more even in academic prose. Furthermore, Biber et al. (1999: 464) have found that some verbs behave idiosyncratically, for instance activity verbs that "involve short-term events without long-term results", such as *glance*, *kiss* or *scream*, rarely co-occur with the PrPf.

Previous research has found that stative verbs more often co-occur with the SPst than with the PrPf (Elsness 1997: 321), which in turn suggests a relatively low frequency of stative verbs with the PrPf, an aspect to be investigated for the current data. In addition, studies on language acquisition have hypothesized that L2 learners of English are able to adequately use the perfect with verbs in descending order of frequency on the following hierarchy of Aktionsart categories: states > activities > accomplishments > achievements (Liszka 2003: 18). Provided that learner effects play a role, this suggests that the present L2 data should reveal higher proportions of state verbs and lower values for achievement verbs when compared to L1 varieties, while activities and accomplishments should cover the middle ground. It has furthermore been suggested that, due to substrate influence, some L2 varieties are generally indifferent as regards Aktionsart (Kachru & Smith 2008: 90–91). This would imply that idiosyncratic combinations of continuative semantics and accomplishment verbs (for instance) occur, while it has to be acknowledged that this is probably more relevant for the extended range of usage of progressive forms that is conspicuous in L2 varieties of English (cf. eWAVE Feature 88/89 2012).

4.3.3 Sentence type

Another contextual factor for which the present dataset was coded is the sentence type in which PrPf forms occur. Earlier research has found that PrPfs are favored in negative sentences (Elsness 1997: 321) and that the use or non-use of temporal adverbials is sensitive to sentence type (Schlüter 2002a: 242). Additionally, the semantic interpretation may in some cases be determined by this factor (Winford 1993: 166). In light of these findings, it was deemed worthwhile to include the coding for sentence type. A straightforward threefold coding scheme was applied:

(i) Positive statements (pos; e.g. *I've been in this limbo for too long*; ICE-CAN s1a-037)
(ii) Negative statements (neg; e.g. *Their performance has not been good*; ICE-NZ w2c-014)

(iii) Questions (ques; e.g. *How many have locked themselves up in their own room*; ICE-PHI s1b-018)

4.3.4 Semantics

As the extended review of the literature on the semantics of the PrPf has shown (see Section 3.3), innumerable approaches and in consequence potential coding schemes are available, many of which suffer from inherent weaknesses due to their having been arrived at by mere introspection. Therefore, in order to minimize the amount of subjectivity involved in establishing the coding scheme and for reasons of comparability, I follow Schlüter's (2002a: 150–171) fourfold model devised on corpus-based principles.[224] It distinguishes the following categories:

(i) Indefinite past – single acts/events (indef-s; e.g. *the teacher who taught us Bridge has gone*; ICE-HK s1a-042)
(ii) Indefinite past – multiple acts/events (indef-m; e.g. *I've probably seen them a dozen times*; ICE-AUS s1a-071)
(iii) Continuative past – continuous acts/events (cont-c; e.g. *Since 1988 her bouts of binge-vomiting have grown sporadic.*; ICE-IRL w2b-024)
(iv) Continuative past – states (cont-s; e.g. *The economic team of the present administration during its first year has been superior in cohesiveness effectiveness and clarity in purpose*; ICE-PHI s2b-026)

It was outlined in the review of previous corpus studies that some of these semantic meanings are more central (e.g. single act indefinite pasts) while others are less so (e.g. continuative state pasts); that is, they are relatively more or less frequent.[225] With the help of the findings

[224] For a detailed description of Schlüter's (2002a) model and the criteria and selection process he applies to establish diverse semantic readings, see Section 3.5.3.
[225] Cf. also Klein & Vater, who distinguish between a *basic* meaning that "can be obtained by abstraction from individual meanings" and a *central* meaning that "is the most frequent one" (1998: 216). This has repercussions for the theory of the PrPf insofar as monosemous accounts usually try to define the semantics of the PrPf in terms of basic meaning whereas

obtained for the present data, we can determine the "centrality" of these readings for the samples of the individual ICE components and furthermore across text types and genres (see Section 7.4). Further, we can compare the values for the semantic categories both among themselves and with the values established in earlier work. This enables us to decide whether a central meaning of the PrPf can be established across all varieties under investigation or whether varieties of different types and locations diverge substantially. In addition, co-occurrence patterns of the individual semantic categories with temporal adverbials can be investigated.

4.3.5 Preceding time reference

Yet another contextual factor that plays a role with respect to the PrPf is preceding tense (or, in a broader view, preceding forms that create temporal reference; henceforth TR forms). Earlier research has noted that PrPf contexts are preceded by a present tense in the vast majority of cases (Davydova 2011: 123) and that with regard to HAVE-perfects co-occurring with a definite temporal adverbial in particular, other perfect forms or present tense forms often appear in the preceding context, supposedly acting as triggers for the PrPf (Davydova 2011: 157). The subsequent categories, also distinguishing between different aspectual forms, were used:

(i) Simple Present (PrTs)
(ii) Present Progressive (PrTp)
(iii) Simple Past Tense (SPst)
(iv) Progressive Past Tense (Pstp)

---

polysemous accounts allow central and less central meanings. Proponents of monosemous PrPf theories thus have to search for a "lowest common denominator" such as anteriority or current relevance, which may be vague and relatively imprecise. In contrast, polysemous accounts in principle allow for more flexibility, although this might imply arbitrariness when done by mere introspection. It could be argued that polysemous approaches, ideally not simply employing seemingly "universally recognized" preconceived semantic categories, are more suitable for quantitative study, while the search for a central meaning does not preclude the deduction of a basic meaning in principle.

(v) Simple Present Perfect (SPrPf)
(vi) Present Perfect Progressive (PrPfp)
(vii) Past Perfect (PaPfs)
(viii) Past Perfect Progressive (PaPfp)
(ix) Forms with futurate reference (F)

Note that formulaic utterances typical of spoken registers such as *I mean* or *you know* (see (14) below) were ignored, as they invariantly occur in the Simple Present and thus might inadvertently skew the data (Tagliamonte 2006: 90–91). The same applies for false starts/self-corrections (see also example (14)). If a PrPf occurred within the verbal element of a subordinate clause, the tense of the main verbal form of the superordinate clause was coded (see (15)), while if the PrPf form was a main verb itself, as in (16), the tense of the main verb of the preceding sentence or utterance was coded.

(14) So it's it's *I mean I think* that you'll find that although the pendulum at the moment has swung completely over to the side of uh local treatment for cancer of the breast (ICE-GB s1b-010; coded as F)
(15) I *do not recollect* whether I have stated before (ICE-IND s1b-064; coded as PrTs)
(16) Further testing *was* too costly so I did not pursue this course. (Unfortunately I have misplaced the laboratory results also.) (ICE-AUS w1b-098; coded as SPst)

Again, the coding for preceding tense facilitates the establishment of co-occurrence patterns of this factor with PrPf forms across a number of varieties of different types as well as across various text types and genres, so that typical temporal environments for PrPf constructions and potential priming effects can be identified. It also allows us to compare the findings for the ICE components with hypotheses from earlier research, and thus serves to complement these.

4.4 Multidimensional aggregational analysis

Statistical techniques such as distributional tables and corresponding tests for the significance of differences, correlations and standard deviation (SD) values to assess the dispersion of data have long been established in linguistic analysis. In addition to these measures, the current study also employs multidimensional aggregational analysis as an exploratory method to establish measures of similarity across varieties, genres and text types. Two types are applied: hierarchical cluster analysis and non-hierarchical phylogenetic networks.

While cluster analysis has frequently been used as a tool for creating and testing hypotheses in other fields of research, such as natural and other social sciences (see e.g. Romesburg 1984), its application in linguistics in general and in the study of World Englishes in particular as yet has been restricted (but see Manning & Schütze 1999: 495–528 for an outline of its potential applications in linguistics and Gries & Stefanowitsch 2010 for a pertinent example).[226] Thus, a few remarks on this statistical method are in order.

The usefulness of cluster analysis lies in the fact that it enables the researcher to identify a latent structure that is not directly accessible with the help of manual analysis alone in extended sets of multidimensional data material consisting of a large number of individual data points. Its main aim is to graphically represent and reveal relationships of similarity and dissimilarity between different items, with unrooted tree-shaped dendrograms as the final graphical output. To achieve this, cluster analysis compares items pairwise within a similarity matrix and then fuses all items into clusters that are (depending on the clustering method used) internally maximally similar or minimally dissimilar but, in any case, highly dissimilar to other clusters and items (Manning & Schütze 1999: 501).

Therefore, hierarchical cluster analysis represents a statistical method that can be employed to adequately establish a measure of simi-

[226] Szmrecsanyi & Kortmann (2009) also establish clusters of vernaculars of English according to variety type; however, they use an alternative method, principal component analysis.

larity or dissimilarity[227] between the items under investigation in the present study, such as different varieties, genres, or text types that are coded for the factors described in the preceding sections. These items cover the different leaves in the graphs. In addition, cluster analysis provides a bird's eye perspective of a large dataset, which allows the findings for the PrPf to be related to a number of the general models of World Englishes as outlined in Section 2.5.1 and the adequacy of these models to be determined for the grammatical area under investigation. As it further provides a measurement of the distance between varieties, the claim made in earlier corpus-based studies (e.g. Hundt & Vogel 2011) can be tested, namely that, when analyzing individual grammatical features, varieties are unexpectedly congruent irrespective of their type as defined above (Section 2.5.1). This, in turn, would suggest that varieties do not differ substantially in what are conceived as "core" structural areas.

In the present study, the absolute values obtained through the coding of the individual factors are translated into relative values.[228] To make the data readable for the calculation of the similarity matrices and further processing, these values are entered into comma-separated files (CSV files), where each column represents a register, macro-genre, or text type, and each line contains the values of the same category that are to be compared.[229] The hierarchical cluster analysis based on the similarity matrix then provides a graphical output in the form of a tree-like dendrogram and a measure of how well the graphical representation

[227] This approach modifies one of the possible approaches discussed by Davydova et al., namely establishing "proximity of a variety to some reference variety" (2011: 294). In cluster analysis all varieties serve as reference varieties for each other as no prevalence is given to one or the other variety as a point of reference or "gold standard".

[228] These relative values are the ones presented in Chapter 5 below for each of the varieties and the respective registers, macro-genres, and text types (unless stated otherwise). Thus, potential skewing effects of the slightly varying sample sizes of the individual corpus components (see Section 4.2) could be avoided. All data points had values between 0 and 1 so no standardization was needed. Standardization (e.g. mean-based standardization; Moisl 2010) could be imperative, however, when using absolute values in cluster analysis.

[229] To calculate statistical measures, R was used. For clustering (function *hclust*) I employed the complete linkage method (Manning & Schütze 1999: 505–507) that is based on the similarity of the two least similar members of a cluster. Rank-based distance matrices were created with the Spearman method (Ghada Mohamed, University of Lancaster, p.c.).

actually fits the corresponding numerical data from the distance matrix (see further Section 4.5).

As an alternative (non-hierarchical) means of depicting similarities across varieties, I employ phylogenetic networks or phenograms that also rely on similarity matrices as input. Originating from the area of evolutionary biology (Huson & Bryant 2006),[230] in linguistics these have been used mainly in historical and typological studies, and in English linguistics in particular in the area of dialectometry (e.g. Szmrecsanyi & Wolk 2011), large-scale comparisons of morphosyntactic profiles (e.g. Kortmann & Wolk 2012), as well as for the comparison of lexical and accent differences across varieties (McMahon & McMahon 2005; McMahon & Maguire 2013; cf. Wichmann & Urban 2012).[231] They produce an unrooted network representation (NeighborNet) that establishes, first of all, "geolinguistic signal[s]" (Szmrecsanyi 2012) in the data. These network representations then allow for a more fine-grained analysis. Differences, in our case between varieties and text types, are not merely shown in terms of absolute cluster membership that is branched categorically and always bifurcating or neighbor-joining, but rather in terms of relative distances to *each* of the other categories (varieties and text types) that cover the terminal nodes in the graphs. Boxes (splits) in the NeighborNet output can be interpreted as representing contact or at least interrelation whenever two nodes share such a split. Furthermore, this method of graphical representation enables us to assess differences between categories that would form members of a single cluster in a cluster analysis, in other words between those that are, e.g., minimally dissimilar and therefore cluster together.

4.5 Caveats

It should be noted that any type of multidimensional analysis involves a certain amount of abstraction from linguistic reality: first, as regards the

[230] For the creation of the NeighborNet representations I used *SplitsTree* 4.12.3 ("equal angle" method). The most recent version of the program is available at www.splitstree.org. As input it requires a nexus file that contains a similarity matrix (see above) created in R.
[231] For further linguistic applications see Nichols & Warnow (2008).

choice and type of numerical values that are entered into the similarity matrix and second, as regards the fit of the graphical representations in relation to the actual measure of similarity that is inherent in the numerical values of the tabulated data.

The former aspect requires awareness on the part of the researcher, as there are a potentially infinite number of language-internal variables that could be coded. However, not all of them are equally meaningful. Therefore, to facilitate coding for the present study, variables were chosen (i) that have been identified in the previous literature as influencing the usage of the PrPf and (ii) for which long-established und universally accepted (or corpus-derived) categorizations are available. This does not preclude the consideration of further contextual factors in the future, should the need arise to incorporate them.

The fit of the matrices and the graphical representations can be assessed for hierarchical cluster analysis with the help of the cophenetic correlation coefficient *c* that is calculated automatically.[232] Numbers are provided with each dendrogram accordingly. *SplitsTree* also provides a value that assesses the fit for the NeighborNet graphs. Due to the fact that more subtleties in the actual data structure can be represented in these graphs, fit values without exception approximated the optimal values and are thus not provided for each relevant figure individually.

Two minor issues are worth noting. Nichols & Warnow (2008: 779) have pointed out that, ideally, results obtained through phylogenetic methods should be compared with a benchmark dataset (which is established e.g. through qualitative historical evidence), as statistical means of assessing the significance of splits (e.g. through bootstrapping, see below) still need to become more sophisticated. However, we have to con-

[232] *C* takes values between 0 and 1. The closer the value is to 1, the better the fit between the graphical representation of the clusters and the actual pairwise distance between data points in the original numerical data (Ghada Mohamed, University of Lancaster, p.c.; see further Sokal & Rohlf 1962). As regards levels of cophenetic correlation values, Romesburg (1984: 27) states that there is no universally agreed categorization but that c-values of 0.8 or higher indicate that the distortion inherent in the hierarchical cluster analysis is tolerable. Even though the vast majority of hierarchical cluster analyses in the present study reach c-values that are well above the 0.8 threshold, these representations are always complemented by non-hierarchical analyses to compensate for potential skewing effects and to obtain a more meaningful picture.

tend with the situation that no comparable benchmark dataset is available. Partial remedies to this could be comparison with models of World English, which also try to depict similarities and differences between varieties, as well as cross-references to findings from studies with a related focus whenever applicable and cross-validation of the findings with different types of data, for instance questionnaire data (Krug & Sell 2013). Furthermore, Sand (1999: 121) has drawn attention to the notorious problem of inconsistencies in transcribed spoken corpus material. For the present study this has ramifications in particular at the tagging stage (see Section 4.2) but may also have consequences for individual corpus searches (such as for the identification of alternative surface forms; see Section 6.4). However, for the present analysis, which aims at providing a global picture, the benefit of comparability across these categories clearly outweighs the influence that individual inconsistencies in transcription (also due to the many teams of researchers that were involved in the compilation of ICE) might have on the overall outcome.

Therefore, while these caveats have to be kept in mind, the potential insights from the analysis make an empirical comparison seem far superior to refraining from it altogether due to minor methodological scruples. In the end, the empirical results will have to be checked for consistency and theoretical plausibility and, where necessary, subjected to re-evaluation or re-analysis with different methodologies or based on different datasets (cf. Krug et al. 2013 on the empirical cycle, principles of empiricism and methodological pluralism).

5 The Present Perfect in worldwide varieties of English

This chapter aims to provide a quantitative and qualitative analysis of the PrPf in the various varieties of English based mainly on language-internal contextual information (Klein 1992: 538–539; cf. Portner 2003: 475) and to establish potential further connections to language-external factors (such as variety and text types). While the present chapter is mainly descriptive and reports findings, Chapter 6 below will contextualize the results and discuss their wider implications.

In general, the analysis follows three lines of investigation. First, the sociohistorical context for each variety (apart from BrE; see e.g. Baugh & Cable 2013 for a concise overview) is briefly outlined and general remarks about the PrPf in the respective variety, as reported in the extant literature, are given. In addition, the categorization of the varieties according to Schneider's (2007) dynamic model is indicated for each variety (see also Table 5.1 for an overview). Second, quantitative data for the individual factors and register effects are considered and illustrated with qualitative data if required. This is followed by a section that assesses the PrPf-friendliness and the congruence of text types in each variety by way of a multidimensional analysis.

Table 5.1. Position of the varieties under investigation in Schneider's (2007) dynamic model

phase	**3 nativization**	**4 endonormative stabilization**	**5 differentiation**
variety	HKE IndE PhiE →	JamE SinE	CanE IrE NZE AusE

5.1 British English

Despite its arguable loss of status in the face of rising influence of AmE and further varieties such as AusE and NZE, BrE still is often used as reference variety, that is, a kind of yardstick against which other varieties and their potential deviations are measured (see e.g. Davydova 2011). At the same time, BrE represents the variety where the PrPf has been studied most widely, often in contrast to AmE. Here, conflicting results with regard to the overall frequency of the PrPf (in contrast to the SPst) seem to emerge, with some claiming a higher frequency in BrE (e.g. Leech 2004: 43, Yao & Collins 2012 or the studies listed by Ritz & Engel 2008: 139), while others reject such an analysis on the basis of their data (e.g. Wynne 2000: 46; see further Section 3.5 above). Ahlqvist has speculated that a lower frequency of the PrPf may be due to the large number of Irish immigrants to the US which did not use the PrPf (2000: 179), while others state that the extent of grammaticalization of the PrPf in the respective varieties is responsible for differences in frequency distribution and usage (e.g. Hundt 1998: 74; Tagliamonte 1999: 377). However, the view that PrPfs claim more and more territory from the SPst in BrE, as maintained by Sempere-Martinez (2008: 124), clearly constitutes a minority position.

5.1.1 Temporal adverbials

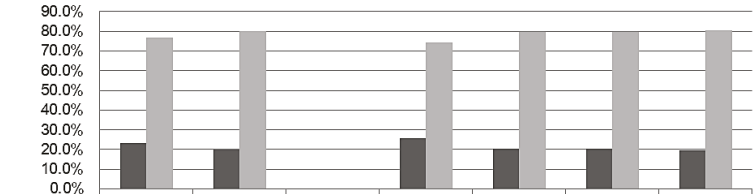
Table 5.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s)

	Ota 1963	Elsness 1997	Wynne 2000	Schlüter 2002a	Hundt & Smith 2009	van Rooy 2009	Davydova 2011	ICE-GB
with TA	31%	33%	32%	33%	29-45%	20%	28%	22%
without TA	69%	67%	68%	67%	55-61%	80%	72%	78%

Early quantitative work (Ota 1963: 43) has revealed that the proportion of PrPfs that are temporally specified is comparatively high (31%) in contrast to other forms, such as the PrTs (c. 2%) or the SPst (c. 9%). As

shown in Table 5.2, by and large, this finding has been confirmed in more recent studies (mainly on BrE and AmE), with proportions of temporally specified PrPfs varying between 20 and 45%.

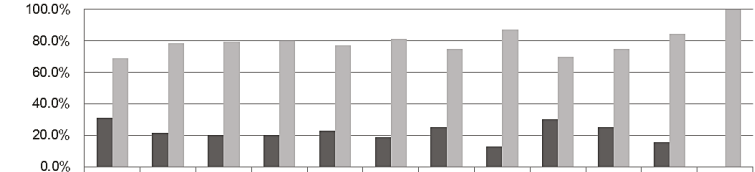
Lower values for specified occurrences of the PrPf can be observed in the studies that also included non-L1 varieties of English (van Rooy 2009 and Davydova 2011), and comparatively lower values of specified PrPf occurrences are also found across the set of ten ICE components (24% in the spoken and 22% in the written parts). Differences between modes of discourse are marginal in the present data, so that exclusive or predominant reliance on written material (as, for instance, in Hundt & Smith 2009) can be disregarded. Therefore, there are two possible explanations for the discrepancy in values between studies that include non-L1 varieties and those that do not: Either a generally lower rate of PrPf occurrences that are temporally specified in non-L1 varieties is evident, which would be surprising as it is also less pronounced and thus only deducible in tendency in Davydova's (2011) data. Low specification values are actually also in evidence in L1 data (ICE-GB), as Table 5.2 shows. Or, more likely, the choice of corpus material in the individual studies exerts some influence. The subsequent analysis of PrPf occurrences with or without temporal specification for BrE and each of the other varieties in the relevant sections provides further evidence for the presence or absence of variety type and mode of discourse/register effects.



	SPOKEN (n=340)	WRITTEN (n=197)		dialogues (n=182)	monologues (n=158)	non-printed writing (n=59)	printed writing (n=138)
with TA	23.2%	19.8%		25.8%	20.3%	20.3%	19.6%
without TA	76.8%	80.2%		74.2%	79.7%	79.7%	80.4%

Figure 5.1.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-GB (register and macro-genre differences)

As can be deduced from Figure 5.1.1, the proportion of PrPf occurrences with temporal specification is slightly, but not significantly[233] higher in the spoken sections of the BrE data. Note further that temporal specification most often occurs in dialogues, which is potentially due to the situational properties of this genre, and least often in printed writing, while none of the distributions deviates significantly from the average overall distribution. Thus, in light of the current data, Miller's statement that "the classic adverb-less PrPf exists only in formal written English" (2004a: 231) seems to lack an empirical basis (see also Section 7.4).



	GB S1A (n=80)	GB S1B (n=102)	GB S2A (n=59)	GB S2B (n=99)	GB W1A (n=22)	GB W1B (n=37)	GB W2A (n=24)	GB W2B (n=31)	GB W2C (n=30)	GB W2D (n=16)	GB W2E (n=26)	GB W2F (n=11)
with TA	31.3%	21.6%	20.3%	20.2%	22.7%	18.9%	25.0%	12.9%	30.0%	25.0%	15.4%	0.0%
without TA	68.8%	78.4%	79.7%	79.8%	77.3%	81.1%	75.0%	87.1%	70.0%	75.0%	84.6%	100.0%

Figure 5.1.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-GB (detailed text categories)

Looking at the individual text types, as shown in Figure 5.1.2, the values for specified PrPfs oscillate between 0% (creative writing; w2f) and 31%

[233] All calculations of significance and effect size are based on the absolute number of occurrences. The function *assocstats* from the library *vcd* was employed to obtain Pearson's Chi-square and Cramer's V (effect size φ_C) values. If one or more cells of a contingency table contained an observed value of five or less, Fisher's exact test was also used. However, in the course of the study, it emerged that the overall results (i.e. significant vs. non-significant differences) more often than not were indifferent to the type of test that was used. This is mainly due to the property that degrees of freedom always were two or more (apart for the analysis of the presence or absence of temporal adverbials, where cell values naturally always were bigger than five) and Pearson's Chi-square in this case also represents a robust test for significance, even if some cells have a value of zero (Larntz 1978: esp. 260–262). Therefore, subsequently only the p-values for Pearson's Chi-square (and the effect size) are reported if a significant difference exists, unless indicated otherwise. For reasons of readability, p-values are presented in relation to the common significance levels of 5%, 1%, 0.5%, and 0.1% (e.g. an exact p-value of 0.00464 is reported as $p < 0.005$) or rounded to the third decimal place when bordering the 5% level.

(private dialogues; s1a). Figure 5.1.2 further shows that in BrE variability is most pronounced within the written categories, where both the lowest and the second highest proportions (reportage; w2c) appear, while values are similar amongst the spoken text types, with one fifth of all instances showing temporal specification by an adverbial (with the exception of private dialogues). Therefore, register and text type effects, as assumed earlier (Panzner 1995: 147), are indeed in evidence here, although they are partly contrary to expectations (cf. Weinrich 2001: 91 on high proportions of specification in narrative texts).

For the distribution of the temporal adverbials in BrE according to the scheme presented in Section 4.3.1, the following picture emerges:

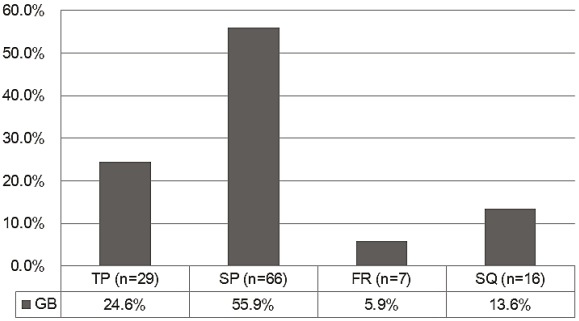
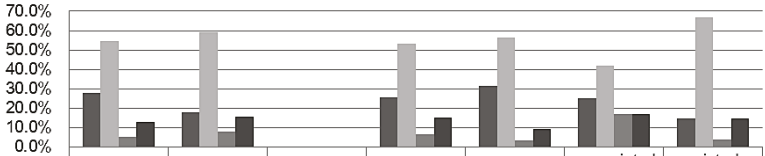


Figure 5.1.3. Types of PrPf adverbial specification: ICE-GB

The most frequent means of temporal specification of PrPf constructions covering more than half of all relevant instances, is by adverbials of span and duration (SP), as in (1), while adverbials of time-position (TP), as in (2), and sequence (SQ), as in (3), cover the middle ground and adverbials of frequency (FR), as in (4) are rare.

(1) Eddie Gibson, who has lived on the estate *for more than 20 years* and is a coordinator of the residents' fight, said everyone was delighted with the decision. (ICE-GB w2c-011)
(2) Well that's what I think perhaps he's perhaps he's changed it *now* but I was there last week [...] (ICE-GB s1a-099)
(3) We've seen a couple of occasions *already* in this game [...] (ICE-GB s2a-015)

(4) I have tried to phone you *a couple of times* this week although not persistently. (ICE-GB w1b-001)[234]



	SPOKEN (n=79)	WRITTEN (n=39)		dialogues (n=47)	monologues (n=32)	non-printed writing (n=12)	printed writing (n=27)
TP	27.8%	17.9%		25.5%	31.3%	25.0%	14.8%
SP	54.4%	59.0%		53.2%	56.3%	41.7%	66.7%
FR	5.1%	7.7%		6.4%	3.1%	16.7%	3.7%
SQ	12.7%	15.4%		14.9%	9.4%	16.7%	14.8%

Figure 5.1.4. Types of PrPf adverbial specification: ICE-GB (register and macro-genre differences)[235]

A closer inspection of register differences, as shown under Figure 5.1.4, reveals that the overall picture is largely reproduced for both modes and the broader ICE genres. SP adverbials are responsible for the largest proportion of temporal adverbials across all textual categories. Non-printed writing seems to constitute an outlier, yielding a comparatively more even distribution across the four adverbial categories, which deviates (although not significantly) from the global average. Note again, that due to the low absolute number of occurrences for some genre categories (e.g. 12 in non-printed writing), these findings have to be treated with caution.

Overall, however, based on the ICE-GB data, we can establish a hierarchy of temporal adverbials that typically co-occur with the PrPf in BrE along the lines SP > TP > SQ > FR, which indicates that in this variety the PrPf, when accompanied by a temporal adverbial, most typically surfaces in durative contexts, while other environments such as with a point in time or sequential reading and those that involve FR adverbials are less common. This finding also answers Miller's question whether

[234] Note that (4) also contains *this week*, an adverbial of the TP type. Only the adverbial that appears first in each sentence/utterance is coded.
[235] Column values in the tables below figures that deviate from 100% are due to rounding.

some FR adverbials (e.g. *ever* and *never*) are "optional additions or, given their frequency in spontaneous speech, are [...] to be analysed as central and obligatory constituents that are occasionally ellipted" (2000: 350). The current data show a low proportion of FR adverbials (also in dialogues) and thus strongly suggest that the latter view can be discarded, while the situation may be different for other varieties (see esp. sections 5.6 to 5.10 below).

5.1.2 Aktionsart

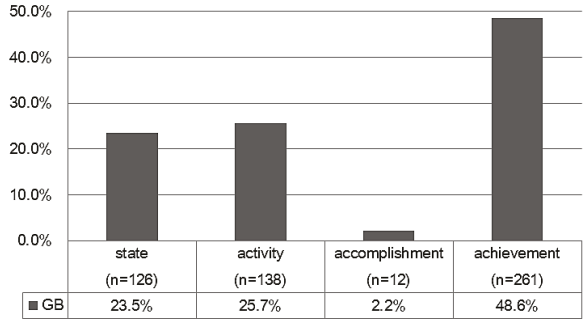


Figure 5.1.5. Aktionsart categories of main verbs in PrPf constructions: ICE-GB

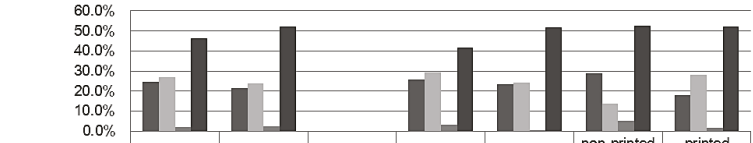
Figure 5.1.5 shows the distribution of the main verbs of PrPfs in BrE according to the Aktionsart categories explained above. Overall, nearly half of all instances involve achievement verbs, as in (5), while one quarter of the main verbs can be categorized as activity verbs, as in (6), and almost one quarter as state verbs, as in (7). Accomplishment verbs, as exemplified in (8), occur only very infrequently.

(5) And concern for the rare birds has *led* to concern for the health of the environment in general, the coral reefs in particular. (ICE-GB w2b-028)

(6) I have *served* in rural and urban areas comprising a variety of political backgrounds (ICE-GB s2b-031)

(7) Honda, which has *had* a collaborative relationship with Rover since 1979, is to build a £ 60 million engine plant there [...] (ICE-GB w2a-015)

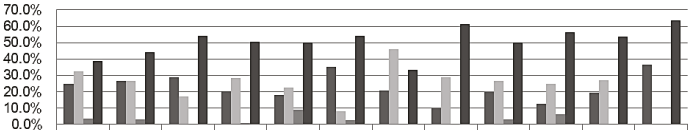
(8) I mean it's assumed if you work for the Royal Shakespeare Company that you've *read the play* or seen the play or even collaborated on a previous production (ICE-GB s1b-023)



	SPOKEN (n=340)	WRITTEN (n=197)		dialogues (n=182)	monologues (n=158)	non-printed writing (n=59)	printed writing (n=138)
state	24.7%	21.3%		25.8%	23.4%	28.8%	18.1%
activity	26.8%	23.9%		29.1%	24.1%	13.6%	28.3%
accomplishment	2.1%	2.5%		3.3%	0.6%	5.1%	1.4%
achievement	46.5%	52.3%		41.8%	51.9%	52.5%	52.2%

Figure 5.1.6. Aktionsart categories of main verbs in PrPf constructions: ICE-GB (register and macro-genre differences)

As regards register differences, Figure 5.1.6 offers a number of insights: Above all, register does not seem to represent an influential factor, as the differences between spoken and written data are not significant. Also in the more detailed view that considers the broad ICE genres, none of the categories shows a significant deviation from the average distribution.



	GB S1A (n=80)	GB S1B (n=102)	GB S2A (n=59)	GB S2B (n=99)	GB W1A (n=22)	GB W1B (n=37)	GB W2A (n=24)	GB W2B (n=31)	GB W2C (n=30)	GB W2D (n=16)	GB W2E (n=26)	GB W2F (n=11)
state	25.0%	26.5%	28.8%	20.2%	18.2%	35.1%	20.8%	9.7%	20.0%	12.5%	19.2%	36.4%
activity	32.5%	26.5%	16.9%	28.3%	22.7%	8.1%	45.8%	29.0%	26.7%	25.0%	26.9%	0.0%
accomplishment	3.8%	2.9%	0.0%	1.0%	9.1%	2.7%	0.0%	0.0%	3.3%	6.3%	0.0%	0.0%
achievement	38.8%	44.1%	54.2%	50.5%	50.0%	54.1%	33.3%	61.3%	50.0%	56.3%	53.8%	63.6%

Figure 5.1.7. Aktionsart categories of main verbs in PrPf constructions: ICE-GB (detailed text categories)

A further look at the individual text types, as displayed in Figure 5.1.7, shows that the rare Aktionsart category of accomplishment verbs is absent from a number of written and spoken text types. It is conspicuous

that the same is true for activity verbs in creative writing (w2f), but this is most likely due to the relatively small random sample that was created for this category (11 occurrences).

Summing up, we can establish a hierarchy of typicality of Aktionsart types that are present in the main verb of the PrPf in BrE along the general line achievement verbs > activity verbs > state verbs > accomplishment verbs, while distributions may vary slightly across text types.

5.1.3 Sentence type

The distribution of sentence types in PrPf contexts in BrE is as follows:

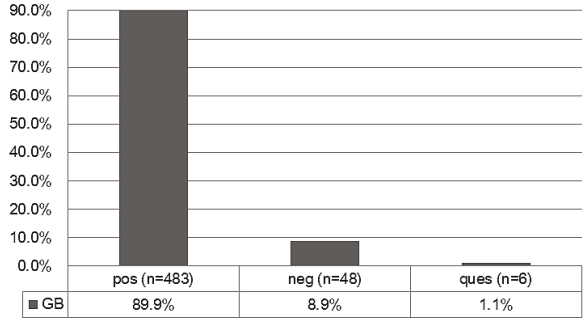
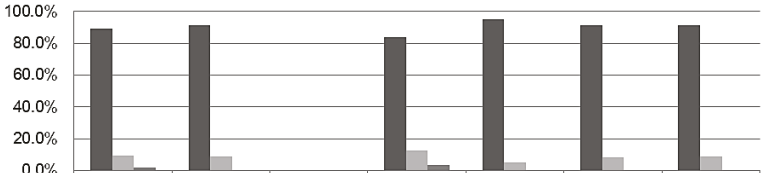


Figure 5.1.8. Distribution of sentence types: ICE-GB

Figure 5.1.8 shows that the vast majority of PrPf occurrences surfaces in positive statements, while less than every tenth instance involves negation (see (9)), and questions with the PrPf, as in (10), are a very rare phenomenon.

(9) And I said oh hi how are you all the rest of it because I haven't spoken for a few weeks because of my work and everything yeah and uh she's not the best correspondent actually you know (ICE-GB s1a-094)

(10) What do you think's gone wrong (ICE-GB s1b-022)



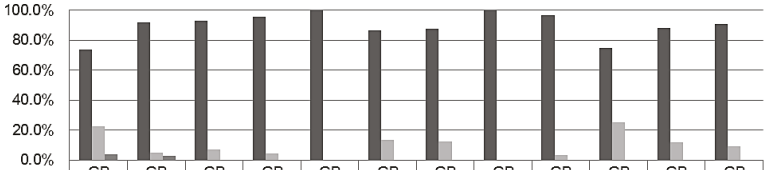
	SPOKEN (n=340)	WRITTEN (n=197)		dialogues (n=182)	monologues (n=158)	non-printed writing (n=59)	printed writing (n=138)
pos	89.1%	91.4%		84.1%	94.9%	91.5%	91.3%
neg	9.1%	8.6%		12.6%	5.1%	8.5%	8.7%
ques	1.8%	0.0%		3.3%	0.0%	0.0%	0.0%

Figure 5.1.9. Distribution of sentence types: ICE-GB (register and macro-genre differences)

With respect to differences according to register and genre, as presented in Figure 5.1.9, positives account for the vast majority of PrPfs in both registers and across all macro-genres. Second, PrPfs in questions are restricted to the spoken part of the data and to dialogues in particular.[236] While questions overall, apart from rhetorical questions, are naturally infrequent in monologues and in some of the written text categories, it is in the dialogues category, with its interactive situational properties, that we would naturally expect most questions overall, and thus higher chances for PrPf questions to appear. Apparently, in the current sample, speakers make use of other grammatical forms when asking for something in the remaining genres where questions and requests would also typically occur (e.g. letters), such as object constructions with verbs of assumption that imply a question, as in (11):

(11) I hope everything has turned out well for you both, and I am sure you've had no trouble with passing your courses. (ICE-GB w1b-008)

[236] The occurrence of the PrPf in questions and the higher proportion of PrPfs in negatives renders dialogues the only macro-genre that differs significantly from the overall distribution of sentence types ($\chi^2 = 6.2827$, df = 2, $p < 0.05$, $\varphi_c = 0.093$), although with a weak effect. In consequence, also the spoken-written contrast is significant ($\chi^2 = 10.319$, df = 2, $p < 0.01$, $\varphi_c = 0.135$). Although relative numbers for negatives vary between 5.1% and 12.6%, none of the other macro-genre categories deviates significantly from the overall distribution.



	GB S1A (n=80)	GB S1B (n=102)	GB S2A (n=59)	GB S2B (n=99)	GB W1A (n=22)	GB W1B (n=37)	GB W2A (n=24)	GB W2B (n=31)	GB W2C (n=30)	GB W2D (n=16)	GB W2E (n=26)	GB W2F (n=11)
pos	73.8%	92.2%	93.2%	96.0%	100.0%	86.5%	87.5%	100.0%	96.7%	75.0%	88.5%	90.9%
neg	22.5%	4.9%	6.8%	4.0%	0.0%	13.5%	12.5%	0.0%	3.3%	25.0%	11.5%	9.1%
ques	3.8%	2.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 5.1.10. Distribution of sentence types: ICE-GB (detailed text categories)

Again, the detailed text categories as presented in Figure 5.1.10 confirm the broad picture. In addition, negatives are fully absent from student and popular writing (w1a and w2b), while they are prominent in instructional writing (w2d). These idiosyncrasies can partly be ascribed to the comparatively low absolute number of occurrences that resulted from the sampling procedure for some of the text categories (e.g. w2d only contains 16, w1a only 22 tokens), which may lead to somewhat less meaningful results in the detailed view. In contrast, the high proportion of negatives in private dialogues (s1a; with more than one fifth of all occurrences of the 80 tokens) suggests a type of discourse that genuinely allows for more negation. Moreover, it involves specification by temporal adverbials much more often than on average (55% of all instances; cf. Table 5.2 above), as illustrated in (12) and (13):

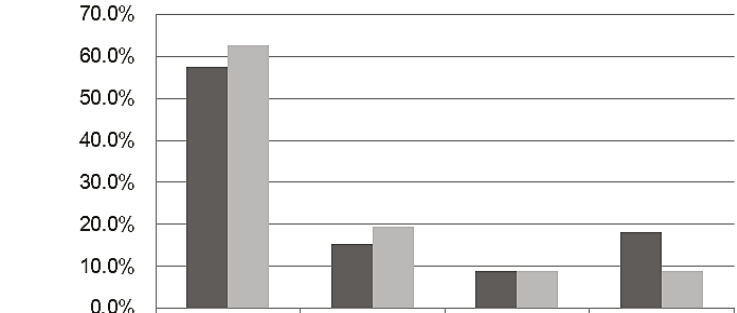
(12) [...] but I had to meet this girl who I haven't seen *for ten years* from my school (ICE-GB s1a-062)
(13) I haven't phoned her *yet* (ICE-GB s1a-100)

Note that the findings for ICE-GB are generally in line with the results of Schlüter (2002a: 135–136), who merely distinguishes between fiction, non-fiction and conversation in BrE (and AmE). His data also suggest that PrPfs occur in affirmative statements in 70–90% of all cases, while negatives in turn are considerably less frequent and questions are very rare. However, the frequency of the last category increases notably in fiction (5.6–13.2%) and in conversation (14%) in his corpora, a finding

that is reproduced in tendency for the comparable macro-genre dialogues but not for fictional texts (w2f) in ICE-GB.

5.1.4 Semantics

Figure 5.1.11 shows the distribution of the four basic semantic categories as outlined above (Section 4.3.4; numbers from Schlüter 2002a are included for reasons of comparison):



	indef-s (n=310)	indef-m (n=82)	cont-c (n=48)	cont-s (n=97)
GB	57.6%	15.2%	8.9%	18.2%
Schlüter 2002	62.8%	19.5%	8.9%	8.8%

Figure 5.1.11. Distribution of semantic categories: ICE-GB vs. Schlüter (2002a: 166)

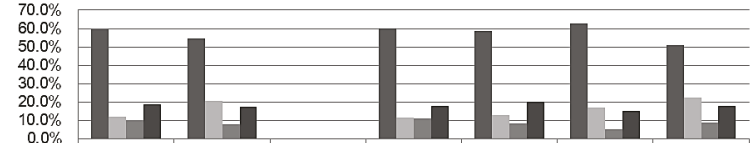
Indefinite single act/event readings (indef-s), as in (14), are present in the absolute majority of cases, while indefinite multiple acts/event readings (indef-m), as in (15), and continuative state readings (cont-s), as in (16), total less than one fifth. The least frequent type is represented by continuative acts/events (cont-c), as illustrated in (17).

(14) It's really embarrassing at the moment because everyone in the street - and I mean everyone - has cut their lawn except us. (ICE-GB w1b-004)

(15) She's had headaches two or three times a week for a couple of hours at a time but she's getting better (ICE-GB s2a-062)

(16) You might say that but I had to meet this girl who I haven't seen for ten years from my school. (ICE-GB s1a-062)
(17) Since January in our authority we've had six hundred referrals of abuse from members of the public other professions (ICE-GB s1b-030)

The distribution in ICE-GB is significantly different ($\chi^2 = 35.757$, df = 3, $p < 0.001$, $\varphi_c = 0.135$) from that found by Schlüter (2002a: 166), which is mainly due to the difference in the continuative states category. In general, however, the finding that indefinite readings overall are highly frequent (four fifths of all occurrences) can be corroborated, while the proportion of continuative acts/events is also strikingly similar across the two sets of data.

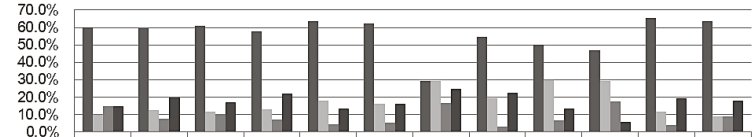


	SPOKEN (n=340)	WRITTEN (n=197)		dialogues (n=182)	monologues (n=158)	non-printed writing (n=59)	printed writing (n=138)
indef-s	59.4%	54.5%		59.9%	58.9%	62.7%	51.1%
indef-m	12.1%	20.7%		11.5%	12.7%	16.9%	22.3%
cont-c	9.7%	7.6%		11.0%	8.2%	5.1%	8.6%
cont-s	18.8%	17.2%		17.6%	20.3%	15.3%	18.0%

Figure 5.1.12. Distribution of semantic categories: ICE-GB (register and macro-genre differences)

Indefinite single acts/events are present in the absolute majority of all occurrences across all macro-genres. The more detailed breakdown in Figure 5.1.12 shows dissimilarities between the two registers due to the higher proportion of indefinite multiple acts/events in the written part. These differences are bordering the 5% level of significance ($\chi^2 = 7.4783$, df = 3, $p = 0.058$, $\varphi_c = 0.1118$). A higher proportion of this category was also in evidence in Schlüter (2002a: 166), who mainly relied on written data, which suggests that register effects play a part here. Note further that both the written and the spoken section are internally homogene-

ous, i.e. their two macro-genres do not differ significantly from each other.



	GB S1A (n=80)	GB S1B (n=102)	GB S2A (n=59)	GB S2B (n=99)	GB W1A (n=22)	GB W1B (n=37)	GB W2A (n=24)	GB W2B (n=31)	GB W2C (n=30)	GB W2D (n=16)	GB W2E (n=26)	GB W2F (n=11)
indef-s	60.0%	59.8%	61.0%	57.6%	63.6%	62.2%	29.2%	54.8%	50.0%	47.1%	65.4%	63.6%
indef-m	10.0%	12.7%	11.9%	13.1%	18.2%	16.2%	29.2%	19.4%	30.0%	29.4%	11.5%	9.1%
cont-c	15.0%	7.8%	10.2%	7.1%	4.5%	5.4%	16.7%	3.2%	6.7%	17.6%	3.8%	9.1%
cont-s	15.0%	19.6%	16.9%	22.2%	13.6%	16.2%	25.0%	22.6%	13.3%	5.9%	19.2%	18.2%

Figure 5.1.13. Distribution of semantic categories: ICE-GB (detailed text categories)

As can be deduced from Figure 5.1.13, indefinite single acts/events are the majority reading for all individual text categories apart from academic writing (w2a), while a number of informational[237] written text categories (academic (w2a) and popular (w2d) writing as well as reportage (w2c)) show a comparatively high proportion of indefinite multiple acts/events, as could be expected from the previous, more coarse-grained analyses. Continuatives are comparatively prominent in academic writing (w2a), particularly in instances where the continuing validity of earlier research or opinions is asserted, as shown in example (16) above as well as in (18) and (19).

(18) Several different types of CCK antagonist have been recognised for some time, but since they have low affinity they have not been suitable for in vivo work. (ICE-GB w2a-027)

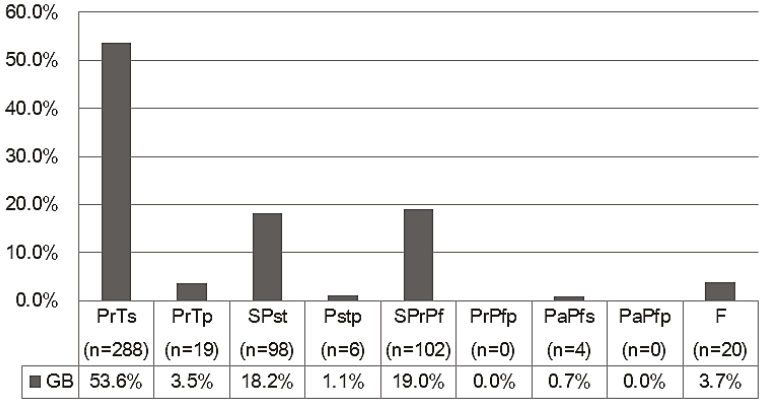
(19) However, lack of telial material on leeks in the U.K. has prevented classification using this system [...] (ICE-GB w2a-028)

[237] Not only can texts be discerned according to the mode of discourse/channel of production (written vs. spoken), but also along other dimensions. For the current study, henceforth I will use a distinction between "involved" and "informational" texts, first established in Biber (1988) as "Dimension 1" along which texts may vary. Applied for the ICE genres, dialogues, monologues and non-printed writing are subsumed under the former, printed writing under the latter category. See Section 4.1 for more detail and discussion.

In addition, the distributions across the four semantic categories of the four spoken text types as indicated by the average SD[238] are much more consistent ($\sigma = 0.02$) when compared to the value for the written categories, which is four times as high ($\sigma = 0.08$) in the ICE data. In sum, this suggests that the distribution is not only susceptible to register but very much dependent on the individual text category. Still, Schlüter's (2002a: 160) finding that indefinite pasts overall are responsible for a share of 80% or more of all instances of the PrPf is corroborated by the present analysis of BrE.

5.1.5 Preceding time reference

For preceding TR forms, earlier research has suggested that PrPfs are likely to be triggered by other PrPfs and after the present (see Section 4.3.5). This observation indeed also applies to the ICE-GB data, as Figure 5.1.14 shows:

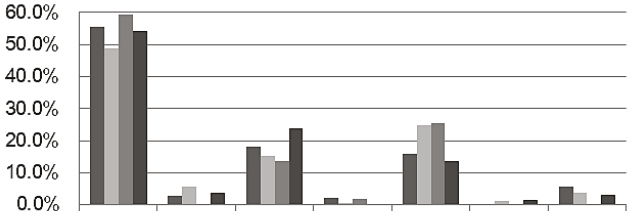


	PrTs (n=288)	PrTp (n=19)	SPst (n=98)	Pstp (n=6)	SPrPf (n=102)	PrPfp (n=0)	PaPfs (n=4)	PaPfp (n=0)	F (n=20)
GB	53.6%	3.5%	18.2%	1.1%	19.0%	0.0%	0.7%	0.0%	3.7%

Figure 5.1.14. TR forms preceding the PrPf: ICE-GB

[238] The average SD is calculated in two steps: First, the SD values are established for each of the four semantic categories for the four spoken and eight written text types respectively. Then, the obtained values are averaged across the four semantic categories. SD values are provided whenever necessary (e.g. for quantifying differences between text types belonging to different modes of discourse).

Figure 5.1.15 provides the breakdown for the macro-genres:[239]



	PrTs (n=288)	PrTp (n=19)	SPst (n=98)	Pstp (n=6)	SPrPf (n=102)	PaPfs (n=4)	F (n=20)
dialogues	55.5%	2.7%	18.1%	2.2%	15.9%	0.0%	5.5%
monologues	48.7%	5.7%	15.2%	0.6%	24.7%	1.3%	3.8%
non-printed writing	59.3%	0.0%	13.6%	1.7%	25.4%	0.0%	0.0%
printed writing	54.3%	3.6%	23.9%	0.0%	13.8%	1.4%	2.9%

Figure 5.1.15. TR forms preceding the PrPf: ICE-GB (macro-genre differences)

It is evident that in each individual macro-genre preceding forms other than PrTs, SPst or SPrPf are rare. While the two spoken genres are overall relatively homogeneous in terms of the proportions of the TR forms, the two written genres differ significantly from each other ($\chi^2 = 12.607$, $df = 6$, $p < 0.05$, $\varphi_C = 0.253$), mainly due to a comparatively high rate (23.9%) of preceding SPst and a corresponding low share (13.8%) of PrPf forms in the printed section. As the proportions of preceding SPrPf versus SPst are almost identical overall and oscillate from genre to genre, further conclusions regarding the influence of the preceding TR form cannot be drawn. Both contexts equally favor PrPfs in the current data, so a priming effect of one single form or the other cannot be postulated here.

[239] As differences between the spoken and the written sections of the data are not significant they are not represented here. No progressive PrPf or PaPf occurred so the respective columns are left out from the figure as well. The same *modus operandi* is applied in the analysis of the remaining varieties.

5.1.6 Perfect-friendliness and text type effects

Earlier corpus-based investigations have found comparatively higher proportions of the PrPf in expository prose in contrast to fictional texts and, more generally, in spoken genres (Schlüter 2002a: 112ff; 2006: 140; see also Bowie et al. 2013). This finding can easily be tested against the ICE data,[240] which yield the following picture:

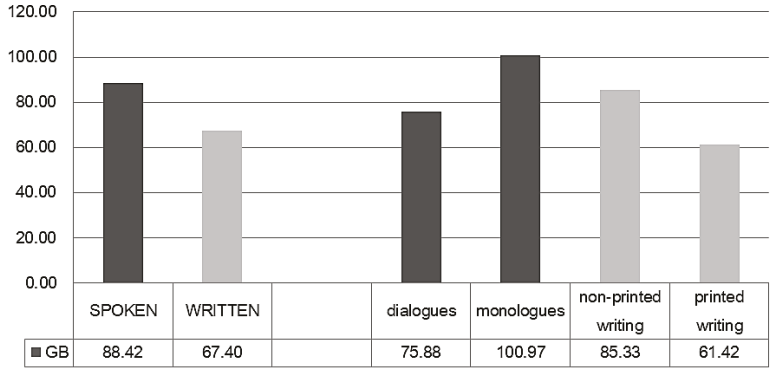


Figure 5.1.16. Frequency of PrPf occurrences in ICE-GB registers and macro-genres (normalized frequency per 10,000 words)

First, while the finding that the PrPf is more frequent in speech (88 PrPfs/pttw)[241] than in writing (67) can be confirmed overall, variability

[240] Rather than a sample (and subsequent extrapolation), the data of the full ICE components (see Section 4.2) is used for the analysis of the overall normalized frequencies of the PrPf throughout Chapter 5.

[241] Other studies (e.g. Yao & Collins 2012) use the PrPf/SPst ratio to assess the PrPf-friendliness of varieties of English. This approach is problematic as it assumes that the PrPf and the SPst are exchangeable one-to-one potentially in all contexts. Schneider & Hundt (2012: 21; but cf. Wynne 2000: 45) have shown in a recent corpus study of IndE and NZE that a high PrPf frequency does usually entail a low SPst frequency while this is not necessarily true vice versa. The result further illustrates that the two TR forms are by no means exchangeable in every case. In addition, other grammatical forms (such as the Simple Present) may be used instead of PrPfs in some varieties (see below and the analyses of variable contexts in Davydova et al. 2011: 300 or Davydova 2011). Therefore, the current analysis relies on normalized frequencies, otherwise known as "Mossé coefficient" (see Hundt & Biewer 2007) as a measure of PrPf salience in the respective varieties.

within each mode of discourse can be observed, with a variance of the normalized frequencies of σ = 29.92 and monologues (101) and non-printed writing (85) showing the highest values.[242] A possible explanation for this finding is that spoken texts favor PrPf usage as this mode of discourse is naturally more self-oriented and in consequence oriented toward current relevance in the widest possible sense (see the discussion of the concept above, Section 3.3.1.2). Therefore, speech may yield a larger proportion of related temporal constructions, such as the PrPf. Furthermore, in spoken texts, and particularly in conversation, speakers share "the same temporal context" (Biber & Conrad 2009: 87) and do not require grammatical forms with definite temporal reference (such as the SPst) as much as in other (e.g. written; Biber & Conrad 2009: 139) genres.

Second, while the value for fictional texts (w2f) is well below the written average (46), Schlüter's (2002a; 2006) result of a particularly low proportion of the PrPf in respect to expository prose (w2a (35), w2b (39) and w2d (40)), is not reflected in the present data. The comparison with data from other ICE components will help to reveal whether this is due to a genuine register or text type effect or represents an idiosyncratic feature of the present set of BrE data. As regards correlations of the normalized frequencies for the individual text categories with other varieties, it emerged that BrE strongly correlates with IrE and IndE (r = 0.86).

A hierarchical cluster analysis (c = 0.94) based on a 12 x 26 cell matrix, which contains the relative values (see Chapter 4) of all the factors discussed in the preceding sections for each text type (see Figure 5.1.17), reveals that the spoken texts align closely in BrE. Private (s1a) and public (s1b) dialogues and unscripted monologues (s2a) even cover neighboring leaves, although a post-hoc test[243] provides no evidence of a significant cluster at the 5% level that contains spoken categories.

[242] See also the results for ICE-GB from Gries (2006), who finds considerable variability in PrPf frequencies (applying yet another measure, namely frequency per number of verb forms in each ICE text type file) across genres and text types, while PrPfs in general are more frequent in speech.

[243] Multiscale bootstrapping, applying the functions *pvclust* and *pvpick* (default settings) from the library *pvclust* in R, was used.

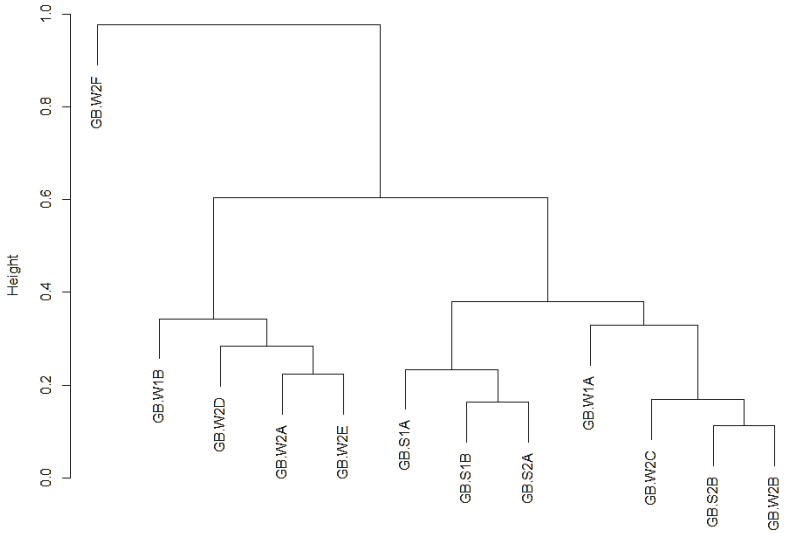


Figure 5.1.17. Cluster dendrogram of similarity across ICE text categories in BrE

That spoken texts align, however, is supported by the non-hierarchical view (see Figure 5.1.18), where short distances between the spoken categories are in evidence. They also share a number of splits.

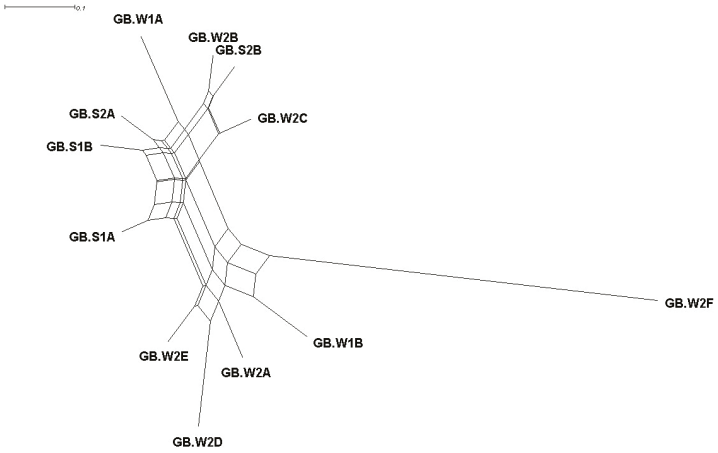


Figure 5.1.18. NeighborNet of similarity across ICE text categories in BrE

In light of earlier register studies this result does not come as a surprise. Biber & Conrad, for instance, although applying a different methodological framework, state that "all spoken texts are surprisingly similar linguistically, regardless of communicative purpose (excluding scripted or memorized texts)" (2009: 261), a claim that nicely tallies with the present data, where the scripted monologue category (s2b) is still close to the other spoken genres in terms of both cluster difference and node distance in the network representation, but does not align directly. In addition, it emerges that creative writing (w2f) covers a structurally conspicuous position in both ways of graphical representation. At the same time, it is striking that even if just one grammatical feature and its related factors are investigated, Biber & Conrad's findings correspond to the present results. This is even more remarkable in light of the commonly held notion that the grammatical areas of tense and aspect are not prone to register influences (Walker 2011: 82). Again, a comparison to the dendrograms and NeighborNets of further varieties will show whether these apparently clear-cut results are observable also there or whether the alignment of text categories of the same mode displays differing patterns.

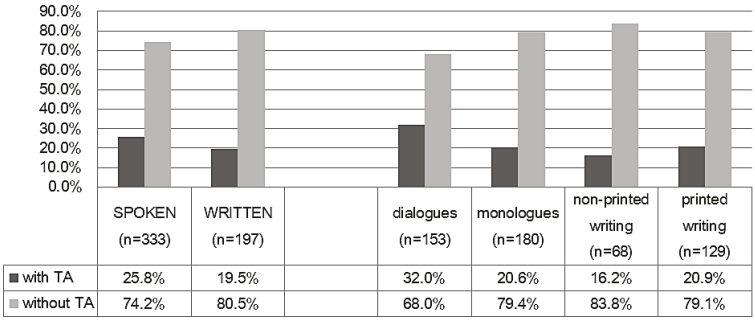
5.2 Irish English (phase 5)

The sociolinguistic situation in Ireland is unique compared to other former colonies of the British Empire: first of all, because of the short geographical distance between Great Britain and Ireland and second, because of the special language policy that was adopted. It has been recognized that the language policy of the Empire was not uniform and that it varied locally from colony to colony, where mostly education of an English-speaking local elite was the focus, while the indigenous language was used by the majority of the population. In contrast, the aim of the language policy in Ireland was to completely replace Irish, the indigenous language (Brutt-Griffler 2002: 29). Indeed, apart from approximately 80,000 speakers of Irish (in the Gaeltacht areas in the West of the island), English, after a considerable period of bilingualism, has now become the native language. This is why IrE is also referred to as a "shift

variety" (Mesthrie & Bhatt 2008: 52; Kortmann & Szmrecsanyi 2011: 281–283), i.e. a variety where a complete shift from one language to another took place over an extended period of time, or even as a member of the "dialects of British English" (Winford 1993: 143; cf. also Leitner 1992: 194).

In the process of shifting, Irish has left its marks on the structure of IrE and arguably also of other varieties of English in the process of colonization (Ahlqvist 2000: 179), which can still be seen today. Besides a distinctive phonology, particular patterns and variants with regard to perfect usage (including alternation with other surface forms; see Section 6.4) have been identified and widely discussed as "Celticised" (Kirk & Kallen 2006: 103) trademark features of this variety (Corrigan 2010; see further Filppula 2012). Therefore, Ahlqvist (2000: 176–178; see also Görlach 1995: 97 and Hickey 2010b: 12) has suggested a reduced frequency of PrPfs in general, while Pietsch (2009: 529) has found a complete absence of the PrPf in basilectal forms. However, the HAVE-perfect is the most frequent variant and can be used in all contexts (Kallen 1989: 21–22; Filppula 2004: 74–76), so whether reduced PrPf-friendliness is the case for the non-vernacular ICE-IRL data is tested below.

5.2.1 Temporal adverbials

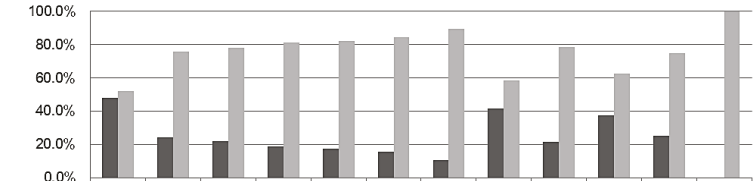


	SPOKEN (n=333)	WRITTEN (n=197)		dialogues (n=153)	monologues (n=180)	non-printed writing (n=68)	printed writing (n=129)
with TA	25.8%	19.5%		32.0%	20.6%	16.2%	20.9%
without TA	74.2%	80.5%		68.0%	79.4%	83.8%	79.1%

Figure 5.2.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-IRL (register and macro-genre differences)

In IrE, nearly every fourth occurrence of the PrPf is temporally specified by an adverbial. Figure 5.2.1 provides the distribution according to regis-

ter and genre. It shows that more specifications occur in the spoken mode of discourse, while the distribution is not significantly different from that in the written section. Again, temporal specification, nearing one third of all instances, most often occurs in dialogues, while the lowest value can be found in non-printed writing. Note that neither of these two genres shows a pattern that is significantly different from the overall distribution.



	IRL S1A (n=50)	IRL S1B (n=103)	IRL S2A (n=95)	IRL S2B (n=85)	IRL W1A (n=17)	IRL W1B (n=51)	IRL W2A (n=19)	IRL W2B (n=36)	IRL W2C (n=28)	IRL W2D (n=16)	IRL W2E (n=20)	IRL W2F (n=10)
with TA	48.0%	24.3%	22.1%	18.8%	17.6%	15.7%	10.5%	41.7%	21.4%	37.5%	25.0%	0.0%
without TA	52.0%	75.7%	77.9%	81.2%	82.4%	84.3%	89.5%	58.3%	78.6%	62.5%	75.0%	100.0%

Figure 5.2.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-IRL (detailed text categories)

A look at individual text types, as presented in Figure 5.2.2, reveals that variability is again more pronounced within the written (specifically the informational) categories. No specification is found in creative writing (w2f) at all, although this finding has to be interpreted with care as only ten instances of the PrPf occur in the present sample. The second lowest value in IrE is in evidence for academic writing (w2a; 24 tokens), while comparatively high values occur for instructional (w2d) and popular writing (w2b), where around two fifths of all PrPfs are further specified. Private dialogues (s1a) are a conspicuous spoken text category with nearly half of all tokens showing temporal specification. Thus, register and text type effects can also be postulated for IrE.

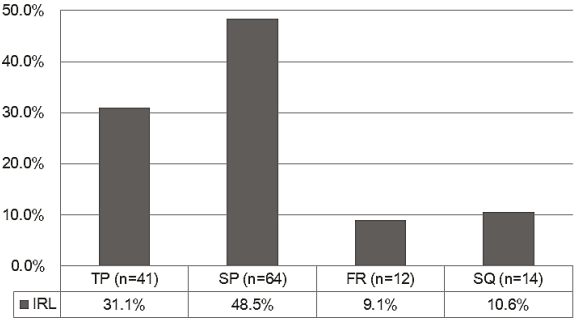


Figure 5.2.3. Types of PrPf adverbial specification: ICE-IRL

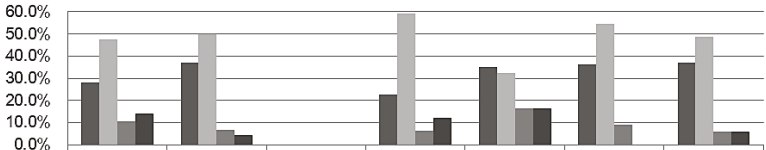
Figure 5.2.3 shows that the major means of specification, which covers almost half of the occurrences under investigation, is by adverbials of span and duration (SP), as in (1). Adverbials of time-position (TP) occur with nearly one third of specified PrPfs, as in (2), while those expressing sequence (SQ), as in (3), and adverbials of frequency (FR), as in (4), are less usual.

(1) I now would agree that all of today's confusion added to the layers of confusion we've had *for the last five weeks* should immediately go to an inquiry (ICE-IRL s1b-032)
(2) Heat the butter, lard and water in a saucepan, and when boiling make sure that the fat has melted *before boiling point is reached*. (ICE-IRL w2d-012)
(3) You *still* haven't answered the question (ICE-IRL s1a-018)
(4) I'm sure you have heard this *on many occasions*. (ICE-IRL w2b-025)

The more detailed picture, as shown in Figure 5.2.4, reveals some variability when contrasting registers and macro-genres. Overall, differences in the distribution between the spoken and written mode are not significant.[244] SP adverbials constitute the main type of temporal specification

[244] Neither does any of the genres deviate significantly from the overall distribution, although the relative frequencies seem to suggest the contrary. This is mainly due to the comparatively low overall absolute numbers that underlie the genre analysis in this partic-

in both registers and all genres apart from monologues, while TP adverbials are also frequent in the written genres. A marked contrast between the two modes of discourse occurs with regard to the low-frequency specifications FR and SQ, which are either completely absent (e.g. SQ in non-printed writing) or yield numbers in the written genres that often amount to no more than half the values of the spoken genres.



	SPOKEN (n=86)	WRITTEN (n=45)		dialogues (n=49)	monologues (n=37)	non-printed writing (n=11)	printed writing (n=34)
TP	27.9%	37.0%		22.4%	35.1%	36.4%	37.1%
SP	47.7%	50.0%		59.2%	32.4%	54.5%	48.6%
FR	10.5%	6.5%		6.1%	16.2%	9.1%	5.7%
SQ	14.0%	4.3%		12.2%	16.2%	0.0%	5.7%

Figure 5.2.4. Types of PrPf adverbial specification: ICE-IRL (register and macro-genre differences)

In sum, based on the ICE-IRL data, we can establish a hierarchy of temporal adverbials that typically co-occur with the PrPf in IrE along the lines SP > TP > SQ > FR. It has to be noted, however, that in contrast to BrE, the frequency of TP is higher than that for SP in one genre, so it can be concluded for IrE that two categories of temporal adverbials, namely SP and, less distinctly, TP, can be seen as typical specifiers of the PrPf.

5.2.2 Aktionsart

Figure 5.2.5 below presents the distribution of the Aktionsart categories for the main verbs in the ICE-IRL dataset:

ular variety (e.g. 11 occurrences for non-printed writing), which makes testing for significance and effect size less meaningful in this case.

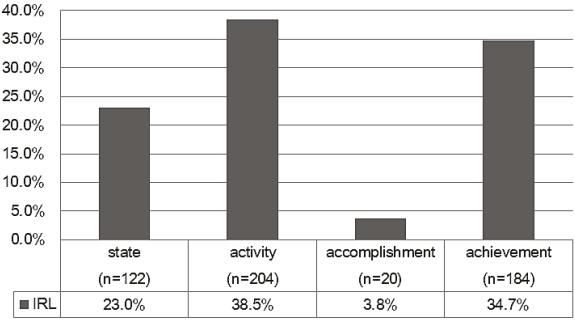


Figure 5.2.5. Aktionsart categories of main verbs in PrPf constructions: ICE-IRL

Approximately two fifths of the occurrences can be categorized as activity verbs as illustrated by (5), while achievement verbs, as in (6), cover more than one third of all instances. Stative verbs, as exemplified by (7), overall amount to more than one fifth, while accomplishment verbs, as in (8), like in BrE represent a marginal class.

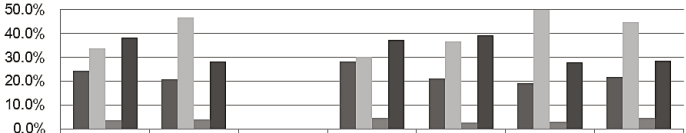
(5) He has *promoted* wellingtons in Cyprus ("it was supposed to be the rainy season but it was blazing sunshine"), played the Mona Lisa in the Louvre, headed the St Patrick's day Parade in Moscow, and was the judge in the Parade of Innocence after the release of the Guildford Four. (ICE-IRL w2c-013)

(6) And I happen to consider that's a very good thing because it has *enriched* the life of the entire community (ICE-IRL s2b-029)

(7) This has *been* the most boring weekend ever. (ICE-IRL w1b-008)

(8) We've managed to read one two three four five six well at least we we we have we *have discussed eight novels* (ICE-IRL s1b-019)

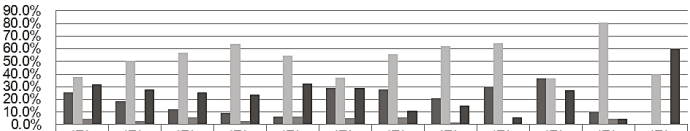
The exact breakdown for registers and genres is illustrated in Figure 5.2.6. It shows that activity verbs in writing and achievement verbs in speech represent the most frequent type.



	SPOKEN (n=333)	WRITTEN (n=197)		dialogues (n=153)	monologues (n=180)	non-printed writing (n=68)	printed writing (n=129)
state	24.3%	20.8%		28.1%	21.1%	19.1%	21.7%
activity	33.6%	46.7%		30.1%	36.7%	50.0%	45.0%
accomplishment	3.6%	4.1%		4.6%	2.8%	2.9%	4.7%
achievement	38.4%	28.4%		37.3%	39.4%	27.9%	28.7%

Figure 5.2.6. Aktionsart categories of main verbs in PrPf constructions: ICE-IRL (register and macro-genre differences)

While none of the genres differs significantly from the overall distribution, the impression that the registers are internally homogeneous is also confirmed by statistical tests. However, when contrasting the aggregated spoken and written sections of the data, significant differences emerge ($\chi^2 = 9.7964$, df = 3, $p < 0.05$, $\varphi_c = 0.136$), mainly due to the higher share of activity verbs and the lower rate of achievement and state verbs in the latter register.



	IRL S1A (n=50)	IRL S1B (n=103)	IRL S2A (n=95)	IRL S2B (n=85)	IRL W1A (n=17)	IRL W1B (n=51)	IRL W2A (n=19)	IRL W2B (n=36)	IRL W2C (n=28)	IRL W2D (n=16)	IRL W2E (n=20)	IRL W2F (n=10)
state	25.8%	18.7%	12.0%	9.3%	6.5%	28.9%	27.8%	20.8%	29.4%	36.4%	9.5%	0.0%
activity	37.9%	50.5%	56.6%	64.0%	54.8%	36.8%	55.6%	62.3%	64.7%	36.4%	81.0%	40.0%
accomplishment	4.5%	2.8%	6.0%	2.7%	6.5%	5.3%	5.6%	1.9%	0.0%	0.0%	4.8%	0.0%
achievement	31.8%	28.0%	25.3%	24.0%	32.3%	28.9%	11.1%	15.1%	5.9%	27.3%	4.8%	60.0%

Figure 5.2.7. Aktionsart categories of main verbs in PrPf constructions: ICE-IRL (detailed text categories)

As could be expected, more variability is in evidence in the detailed view of individual text categories, as presented in Figure 5.2.7. Most text types correspond to the distribution that was established above, insofar that activity verbs are the most frequent class for all written categories (apart from instructional (w2d) and creative (w2f) writing). The rare category accomplishment verbs is absent from the text types reportage (w2c),

instructional (w2d) and creative (w2f) writing, as are state verbs from the last type, probably due to the low absolute number (10) of PrPf occurrences for this text category in the sample.

The detailed view also further illustrates why spoken and written texts differ significantly overall. While the hierarchy of typicality of Aktionsart types that are present in the main verb of the PrPf in IrE all follow the general line activity verbs > achievement verbs > state verbs > accomplishment verbs, distributions vary considerably across the written texts. For instance, state and not achievement verbs represent the second most common type in a group of texts (letters (w1b), academic (w2a) and popular (w2b) writing, reportage (w2c), instructional (w2d) and persuasive (w2e) writing) that are mostly intended for publication. Moreover, some idiosyncratic distributions of the four Aktionsart categories are observable for letters (w1b; high rate of achievement verbs), instructional writing (w2d; high rate of state and achievement verbs), persuasive writing (w2e; almost solely activity verbs) and creative writing (w2f; absence of state and accomplishment verbs). Some of these idiosyncrasies are seemingly attributable to low token counts and a resulting skewed picture. In fact, at least 16 or more instances were coded for seven out of the eight written text types. Therefore, the results are robust and again hint at a particular type of influence different text types may have on the distribution of Aktionsart categories with the PrPf in IrE. In particular, it seems plausible that most texts from the printed writing section of the data employ an extended number of state verbs that support their matter-of-fact style.

5.2.3 Sentence type

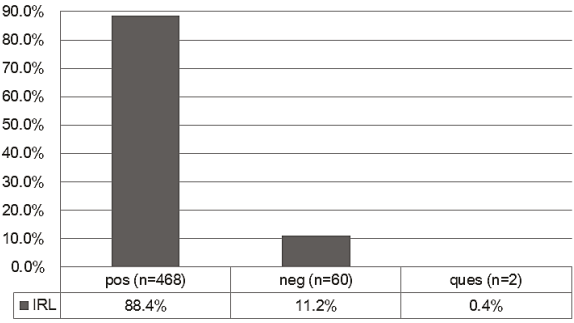
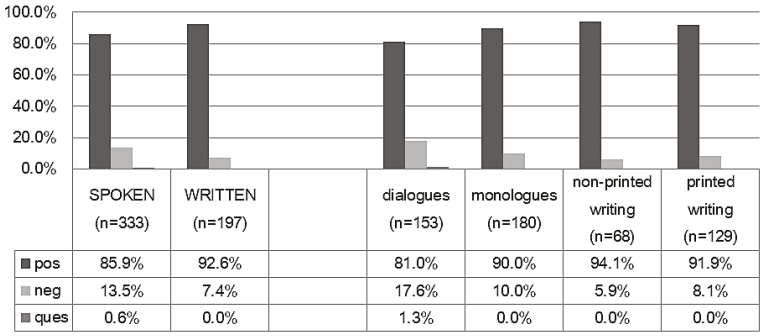


Figure 5.2.8. Distribution of sentence types: ICE-IRL

Comparable to BrE, Figure 5.2.8 illustrates that almost nine out of ten PrPf occurrences surface in positive statements, while every ninth instance involves negation, as in (9). Questions with the PrPf, exemplified in the form of an elliptical tag question in (10), occur again only very sporadically (two occurrences in the ICE-IRL sample).

(9) I haven't really found information about it (ICE-IRL s1b-020)
(10) God he's really had the time of it hasn't he (ICE-IRL s1a-094)

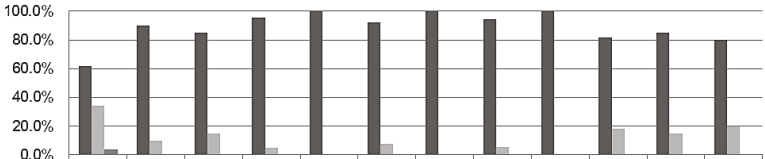


	SPOKEN (n=333)	WRITTEN (n=197)		dialogues (n=153)	monologues (n=180)	non-printed writing (n=68)	printed writing (n=129)
pos	85.9%	92.6%		81.0%	90.0%	94.1%	91.9%
neg	13.5%	7.4%		17.6%	10.0%	5.9%	8.1%
ques	0.6%	0.0%		1.3%	0.0%	0.0%	0.0%

Figure 5.2.9. Distribution of sentence types: ICE-IRL (register and macro-genre differences)

Figure 5.2.9 on register and genre differences reveals some parallels to BrE. First, positives are unmistakably the majority sentence type with

the PrPf across all genres. Second, and more remarkably, dialogues are again the only genre that contains questions, probably due to the interactive situational properties of this particular genre. A third parallel is that negatives are markedly more frequent[245] in dialogues compared to the other genres. The latter two peculiarities result in dialogues being the only genre that deviates significantly from the overall distribution of sentence types ($\chi^2 = 6.5128$, df = 2, $p < 0.05$, $\varphi_C = 0.098$).[246]



	IRL S1A (n=50)	IRL S1B (n=103)	IRL S2A (n=95)	IRL S2B (n=85)	IRL W1A (n=17)	IRL W1B (n=51)	IRL W2A (n=19)	IRL W2B (n=36)	IRL W2C (n=28)	IRL W2D (n=16)	IRL W2E (n=20)	IRL W2F (n=10)
pos	62.0%	90.3%	85.3%	95.3%	100.0%	92.2%	100.0%	94.4%	100.0%	81.8%	85.0%	80.0%
neg	34.0%	9.7%	14.7%	4.7%	0.0%	7.8%	0.0%	5.6%	0.0%	18.2%	15.0%	20.0%
ques	4.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 5.2.10. Distribution of sentence types: ICE-IRL (detailed text categories)

Figure 5.2.10 shows that no negatives were observed in student writing (w1a), academic writing (w2a) and reportage (w2c). In contrast, negatives are very frequent in private dialogues (s1a), representing more than one third of all tokens, which, moreover, are temporally specified in almost half of all instances, as in (11) to (13).

(11) No sure I haven't seen him *since Christmas* (ICE-IRL s1a-003)
(12) Like I haven't visited her *in years* (ICE-IRL s1a-005)
(13) I haven't seen you *in damn ages* (ICE-IRL s1a-098)

[245] Whenever referring to relative frequencies, formulations of the type "variable X is more frequent/common in register/genre/text type Y than register/genre/text type Z" do not preclude that absolute or normalized frequencies of the variable X are actually higher in Z. In other words, these statements refer only to a comparison of the varying proportions of the variable within register/genre/text type Y and Z.

[246] Again, this difference in consequence leads also to the register distinction being significant ($\chi^2 = 6.4342$, df = 2, $p < 0.05$, $\varphi_C = 0.11$), albeit also with a comparatively weak effect in both cases.

This recalls the findings for BrE (see Section 5.1.3). Another text type where the comparatively high proportion of negatives can again be seen as a distinctive feature is instructional writing (w2d).

5.2.4 Semantics

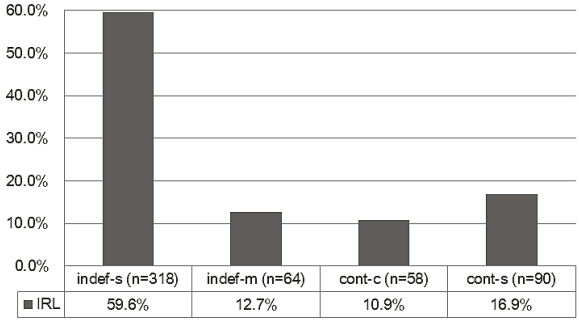


Figure 5.2.11. Distribution of semantic categories: ICE-IRL

Figure 5.2.11 shows the overall distribution of semantic categories for IrE. Indefinite single acts/events (indef-s), as in (14), represent the majority reading, with almost three fifths of all PrPf occurrences. The distribution across the remaining categories indefinite multiple acts/events (indef-m), illustrated in (15), continuative acts/events (cont-c), as in (16) and continuative state readings (cont-s), as in (17), is fairly even, with relative frequencies ranging from 11 to 17%.

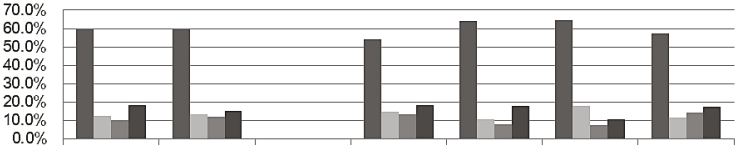
(14) Anthropologist, (sic) George Devereux has stated that "... abortion is an absolutely universal phenomenon." (ICE-IRL w1a-005)

(15) But how Barrett did and I don't know but he came right back and has scored with some good punches every so often particularly with that left hand (ICE-IRL s2a-012)

(16) I've visited those shitty little areas all the time (ICE-IRL s1a-034)

(17) However, in recent years there has been a growing concern among those dealing with radiation matters about the long terms effects of radiation at all levels and the circumstances

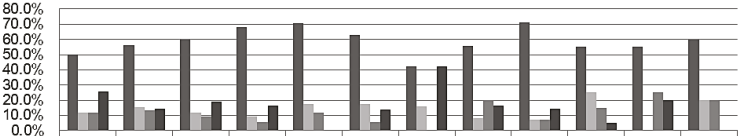
where exposure could, and probably should, be reduced. (ICE-IRL w2b-040)



	SPOKEN (n=333)	WRITTEN (n=197)		dialogues (n=153)	monologues (n=180)	non-printed writing (n=68)	printed writing (n=129)
indef-s	59.5%	59.7%		54.2%	63.9%	64.7%	57.1%
indef-m	12.3%	13.4%		14.4%	10.6%	17.6%	11.3%
cont-c	10.2%	11.9%		13.1%	7.8%	7.4%	14.3%
cont-s	18.0%	14.9%		18.3%	17.8%	10.3%	17.3%

Figure 5.2.12. Distribution of semantic categories: ICE-IRL (register and macro-genre differences)

As is indicated by Figure 5.2.12, indef-s readings are the most prototypical reading across both registers and all macro-genres. While differences between the two modes of discourse are not significant, non-printed writing is prominent as in this genre indefinites of both subtypes are highly frequent and aggregate to more than four fifths of all PrPf instances, while cont-s contexts are very rare. However, this genre is also not significantly different to the overall distribution.



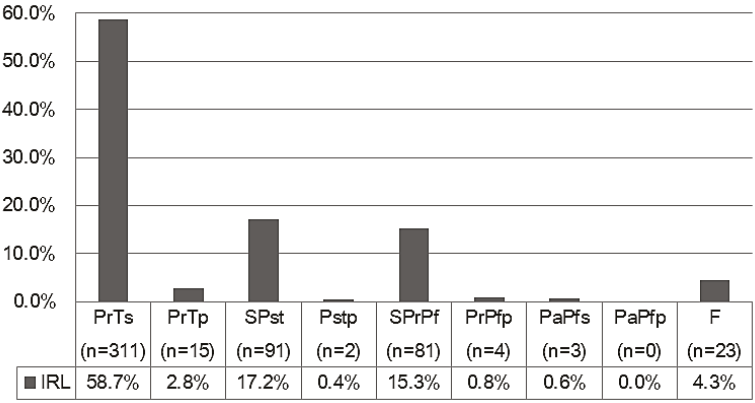
	IRL S1A (n=50)	IRL S1B (n=103)	IRL S2A (n=95)	IRL S2B (n=85)	IRL W1A (n=17)	IRL W1B (n=51)	IRL W2A (n=19)	IRL W2B (n=36)	IRL W2C (n=28)	IRL W2D (n=16)	IRL W2E (n=20)	IRL W2F (n=10)
indef-s	50.0%	56.3%	60.0%	68.2%	70.6%	62.7%	42.1%	55.6%	71.4%	55.0%	55.0%	60.0%
indef-m	12.0%	15.5%	11.6%	9.4%	17.6%	17.6%	15.8%	8.3%	7.1%	25.0%	0.0%	20.0%
cont-c	12.0%	13.6%	9.5%	5.9%	11.8%	5.9%	0.0%	19.4%	7.1%	15.0%	25.0%	20.0%
cont-s	26.0%	14.6%	18.9%	16.5%	0.0%	13.7%	42.1%	16.7%	14.3%	5.0%	20.0%	0.0%

Figure 5.2.13. Distribution of semantic categories: ICE-IRL (detailed text categories)

Figure 5.2.13 displays the breakdown for the individual text categories. It shows that at least half of all tokens can be categorized as indef-s in all

text types except academic writing (w2a). This category seems to constitute a particular case in the ICE-IRL data, as the only one without a single cont-c reading but with a high share of cont-s instead. As in the previous analyses, patterns in the spoken text categories are more comparable with one another, while more outliers can be seen in the written categories,[247] which might have to do with the generally lower number of tokens for the latter in the ICE sample and the potential effect of this on relative frequencies. Therefore, the findings for the written text categories have to be taken with a pinch of salt. Nevertheless, it is evident that the overall distribution of the semantic categories is fairly homogeneous across register and genres, while individual text types may deviate considerably from the general pattern in the current dataset.

5.2.5 Preceding time reference



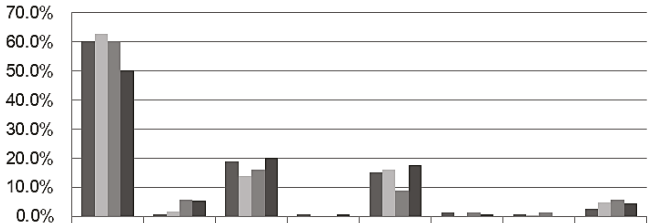
	PrTs (n=311)	PrTp (n=15)	SPst (n=91)	Pstp (n=2)	SPrPf (n=81)	PrPfp (n=4)	PaPfs (n=3)	PaPfp (n=0)	F (n=23)
IRL	58.7%	2.8%	17.2%	0.4%	15.3%	0.8%	0.6%	0.0%	4.3%

Figure 5.2.14. TR forms preceding the PrPf: ICE-IRL

For the distribution of preceding forms expressing time reference (see Figure 5.2.14) the ICE-IRL data yield a similar picture as was found for BrE. The vast majority of PrPfs are preceded by PrTs, followed by preceding SPst, which is the case in almost every sixth token, and SPrPf,

[247] The overall dispersion of the values as indicated by the average SD for the spoken ($\sigma = 0.04$) versus the written text categories ($\sigma = 0.09$) clearly diverges between the texts from the two modes of discourse.

which is the preceding form with more than 15% of all PrPf forms. The remaining forms are considerably less frequent.



	PrTs (n=311)	PrTp (n=15)	SPst (n=91)	Pstp (n=2)	SPrPf (n=81)	PrPfp (n=4)	PaPfs (n=3)	F (n=23)
dialogues	60.1%	0.7%	19.0%	0.7%	15.0%	1.3%	0.7%	2.6%
monologues	62.8%	1.7%	13.9%	0.0%	16.1%	0.0%	0.6%	5.0%
non-printed writing	60.3%	5.9%	16.2%	0.0%	8.8%	1.5%	1.5%	5.9%
printed writing	50.4%	5.4%	20.2%	0.8%	17.8%	0.8%	0.0%	4.7%

Figure 5.2.15. TR forms preceding the PrPf: ICE-IRL (macro-genre differences)

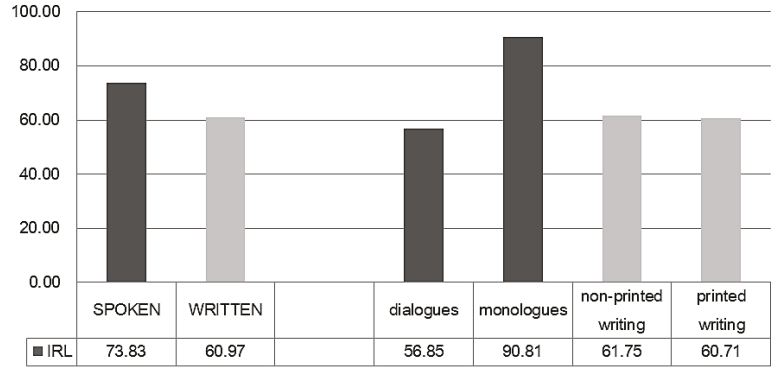
In the macro-genre view, too – as presented in Figure 5.2.15 above – it emerges that preceding forms other than PrTs, SPst or SPrPf are rare. This finding is not surprising given that present or past environments are where we would expect PrPfs to appear. The comparison of the genres shows that the distribution of TR forms in dialogues differs nearly significantly from both the overall pattern ($\chi^2 = 13.980$, df = 7, p = 0.052, $\varphi_C = 0.148$) and clearly significantly from monologues ($\chi^2 = 17.723$, df = 7, $p < 0.05$, $\varphi_C = 0.249$), while the written genres are fairly homogeneous compared both with each other and with the overall distribution.

All in all, this suggests that some genre effects are observable here, although only in the spoken sections. The higher rate of preceding SPst forms is probably due to the purported overall salience of this form (in contrast to the SPrPf) in IrE (cf. Section 5.2).

5.2.6 Perfect-friendliness and text type effects

Figure 5.2.16 shows that IrE, with 65 PrPfs/pttw, is overall on the less PrPf-friendly side, which may be explained by the existence of a range of alternative surface forms that can be used instead of the PrPf in IrE (see

e.g. Winford 2009: 212–214 and Section 6.4). However, more PrPfs again occur in speech (74) than in the written sections (61) of the corpus component. As in BrE, monologues reach a considerably higher value (91) within the spoken genres, while the two written genres align.



	SPOKEN	WRITTEN		dialogues	monologues	non-printed writing	printed writing
IRL	73.83	60.97		56.85	90.81	61.75	60.71

Figure 5.2.16. Frequency of PrPf occurrences in ICE-IRL registers and macro-genres (normalized frequency per 10,000 words)

A closer look at the data reveals that, again like BrE, normalized frequencies are strongly determined by individual ICE text types, for instance with comparatively low values for the categories academic (w2a; 39), popular (w2b; 53), instructional (w2d; 49) and creative (w2f; 22) writing, while the dispersion of the PrPf values amounts to $\sigma = 27.48$. That IrE and BrE behave similarly is further supported by the finding that the normalized frequencies for the individual text categories of these two varieties strongly correlate ($r = 0.86$), although an even higher value is found for JamE ($r = 0.89$).

Figure 5.2.17 presents a hierarchical cluster analysis ($c = 0.92$) for the detailed text categories. In the ICE-IRL data, the picture is less clear-cut than in BrE, with the two monologue categories (s2b, s2b) forming a cluster, while private (s1a) and public (s1b) dialogues align with informational written categories. A post-hoc test reveals that the highlighted cluster is significant. Note that it contains three out of the four spoken categories. However, distances between the clusters that contain speech (and most of the written types) are not very pronounced, as indicated by the small height value.

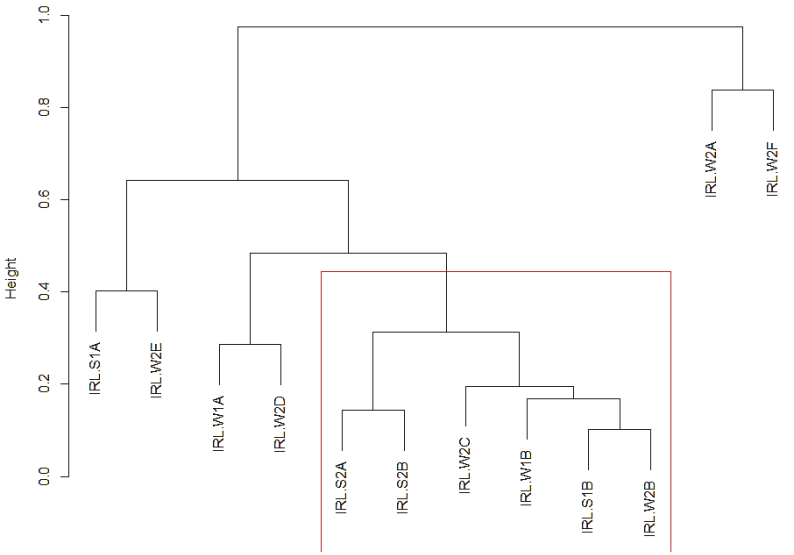


Figure 5.2.17. Cluster dendrogram of similarity across ICE text categories in IrE

This becomes clear in the non-hierarchical NeighborNet, as shown in Figure 5.2.18.

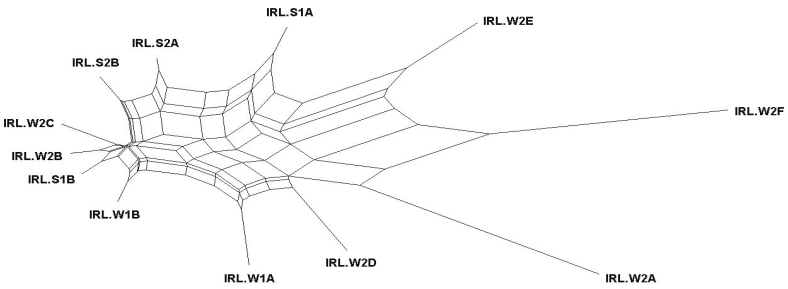


Figure 5.2.18. NeighborNet of similarity across ICE text categories in IrE

In this view, three of the four spoken categories can be found on neighboring nodes toward the top, and the fourth one, public dialogues (s1b), is not overly distant either. Academic (w2a) and creative (w2f) writing, in

contrast, are markedly different from the remaining categories in both types of graphical analysis. Once more, this illustrates the apparent linguistic similarity of spoken and the greater variability of written text categories as postulated by Biber & Conrad (2009: 261).

5.3 Australian English (phase 5)

Australia can be considered a typical example of a settlement (and penal) colony, where immigrants from the British Isles found a new home at the expense of the native Aboriginal population. While the English-speaking white settler group has been dominant since the end of the 18th century, from the middle of the 20th century immigration from other European countries and later also from Asia has been endorsed. In the course of its history, Australia has become more and more independent from its colonial parent, Britain, and this manifests itself both in political (e.g. in the form of the 1986 Australia Act that formalized the comprehensive independence) and linguistic terms. It has been suggested that Australia is now more oriented toward the US, while the country itself constitutes a regional center within the Pacific region due to its size and economic strength.

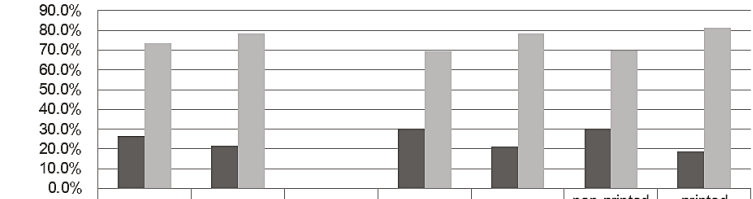
As regards linguistic effects, researchers have hypothesized that due to the relative geographical distance from Britain and the resulting reduced contact between the varieties, AusE can be seen as an independent variety or at least a "semi-center" (Ammon 2005: 1541) of English with characteristic lexical and phonological features. Internally, it is conceived as relatively homogeneous and exerts influence on varieties such as neighboring NZE and allegedly even on BrE (Wächtler 1977: 87–88). It has been shown, however, that diversification into various regional varieties has been ongoing since the 1980s (Schneider 2007: 125–126). At the same time, from a diachronic point of view, Canadian, American (Leitner 1992: 207; Clyne 2005: 299) and Irish varieties of English have left their footprints on AusE.

While a previous study (Fritz 2006: 294) found no particular influence of IrE in the area of PrPf usage (but see Section 6.4), synchronic studies postulate a generalization of the PrPf in AusE, more precisely as

regards the possibility of PrPf use in contexts of co-occurring definite temporal adverbials and its use as a narrative tense. It has been noted above (see Section 3.4) that these findings should be taken with a pinch of salt. Combinations of PrPfs and definite temporal adverbials and narrative use are in evidence in one specialized text type only, which could be interpreted as an indication of the continuing endonormative stabilization of AusE (Collins & Peters 2004: 597–598; see also Ritz 2012: 899). Whether this generalized usage has any repercussions on the distribution of the variables connected to PrPf usage remains to be tested. Still, AusE has been found to be a PrPf-friendly variety in general (Elsness 2009a: 112).

5.3.1 Temporal adverbials

Overall, one in four occurrences of the PrPf is temporally specified in the AusE data. The distributions for speech versus writing and for the individual genres are presented in Figure 5.3.1:

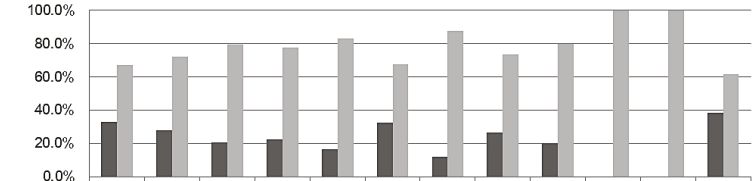


	SPOKEN (n=356)	WRITTEN (n=160)		dialogues (n=198)	monologues (n=159)	non-printed writing (n=42)	printed writing (n=118)
with TA	26.4%	21.7%		30.5%	21.4%	30.2%	18.6%
without TA	73.6%	78.3%		69.5%	78.6%	69.8%	81.4%

Figure 5.3.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-AUS (register and macro-genre differences)

While more specification occurs in the spoken part, the overall picture is fairly homogenous, with the distributions of both modes of discourse not differing significantly from each other or from the overall distribution. Further, none of the genres deviates significantly from the overall distribution or from the values of the respective mode of discourse. Once more, however, there is some tentative evidence that specification

values are determined by genres, as dialogues – the most interactive of the genres – yields the highest rate of temporal specification, exceeding 30% of all instances.



	AUS S1A (n=107)	AUS S1B (n=90)	AUS S2A (n=88)	AUS S2B (n=71)	AUS W1A (n=5)	AUS W1B (n=37)	AUS W2A (n=33)	AUS W2B (n=34)	AUS W2C (n=20)	AUS W2D (n=5)	AUS W2E (n=13)	AUS W2F (n=13)
with TA	32.7%	27.8%	20.5%	22.5%	16.7%	32.4%	12.1%	26.5%	20.0%	0.0%	0.0%	38.5%
without TA	67.3%	72.2%	79.5%	77.5%	83.3%	67.6%	87.9%	73.5%	80.0%	100.0%	100.0%	61.5%

Figure 5.3.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-AUS (detailed text categories)

Figure 5.3.2 shows the results for the individual text types. As in the other varieties, there is considerable variability across these categories despite the relative uniformity that is suggested by the preceding macro-genre analysis. Private dialogues (s1a), at almost one third, have the most specified occurrences among the spoken texts. Variability is even stronger across the written texts. No specification occurs in instructional (w2d) and persuasive (w2e) writing, while nearly two in five instances of the PrPf are temporally specified in creative writing (w2f). Therefore, yet again, text type strongly influences the distributions of temporal specification, although some of the more extreme results might be explained by low token counts (5 tokens for category w2d for the AusE sample).

Figure 5.3.3 shows that the most frequent way of temporal specification of PrPf constructions in the ICE-AUS sample is by adverbials of span and duration (SP), as in (1), exceeding one third of all relevant tokens, while adverbials of time-position (TP), as in (2), cover more than one in four instances. Adverbials of frequency (FR), as in (3), surface with nearly every fourth PrPf, while those of sequence (SQ), as in (4), represent the least common category. However, the latter category still accounts for roughly one in six occurrences that are temporally specified.

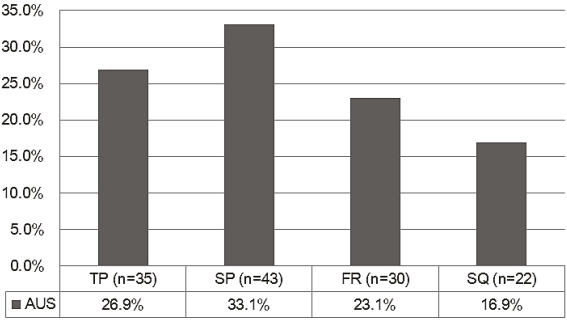


Figure 5.3.3. Types of PrPf adverbial specification: ICE-AUS

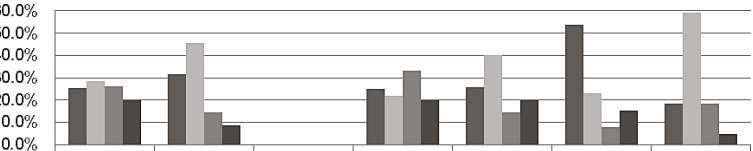
(1) *Since graduating in 1988,* he has worked with both Donaldson + Warn and the BMA, as well as undertaking postgraduate studies as a guest student with Peter Cook at the Stadelschule Hochschule Fur Bildende Kunste in Frankfurt. (ICE-AUS w2a-035)

(2) Mmm but teaching radio to women in the hills is a fairly insignificant thing when we listen back to the the part of the program that you've produced *today* that that you know suggests that that the solution to the new world order is in fact just to redistribute the wealth [...] (ICE-AUS s1b-047)

(3) No she's staying with Jean who's a distant relative who I don't really know and I think I've met *once* (ICE-AUS s1a-091)

(4) As I have said *previously*, some current Vickers distributors will defect because they see us as opposition and they will take on and sell other brand product. (ICE-AUS w1b-030)

It is immediately obvious from Figure 5.3.4 that the mode of discourse and macro-genre distinctions influence the distribution between the adverbial categories in the Australian data. Yet, the apparent difference between speech and writing does not reach the 5% level of statistical significance ($\chi^2 = 4.4951$, df = 3, p = 0.213), while SP in relative terms clearly constitutes the majority variant in writing, and the distribution between the four categories is relatively balanced in speech.



	SPOKEN (n=95)	WRITTEN (n=35)		dialogues (n=60)	monologues (n=35)	non-printed writing (n=13)	printed writing (n=22)
TP	25.3%	31.4%		25.0%	25.7%	53.8%	18.2%
SP	28.4%	45.7%		21.7%	40.0%	23.1%	59.1%
FR	26.3%	14.3%		33.3%	14.3%	7.7%	18.2%
SQ	20.0%	8.6%		20.0%	20.0%	15.4%	4.5%

Figure 5.3.4. Types of PrPf adverbial specification: ICE-AUS (register and macro-genre differences)

A similar outcome can be seen for differences between the individual macro-genres and both the overall distribution and the distributions for each register.[248] However, relative proportions of categories fluctuate markedly according to macro-genre; for instance, with SP as the most common variant in monologues and printed writing, FR in dialogues and TP in non-printed writing.

Overall, when an occurrence of the PrPf is temporally specified by an adverbial, we can postulate the following hierarchy for AusE: SP > TP > FR > SQ. As in a number of other varieties, this means that temporally specified PrPfs in AusE prototypically occur in durative environments. The other categories are still responsible for more than two thirds of the tokens in the sample and numbers vary markedly, though not significantly, across different registers and genres.

[248] Only between the two written genres can we observe a difference that reaches the 5% level ($p = 0.050$). Here we have one of the rare cases where the p-value of the Pearson Chi-square test diverges ($\chi^2 = 7.3749$, $df = 3$, $p = 0.061$, $\varphi_c = 0.459$) and fails to provide an unequivocally meaningful result in terms of determining the significance of the differences in the two distributions (and Fisher's exact test was used) as five out of the eight cells have a value below five. However, note the strong effect size.

5.3.2 Aktionsart

Figure 5.3.5 displays the global distribution of the Aktionsart categories in AusE:

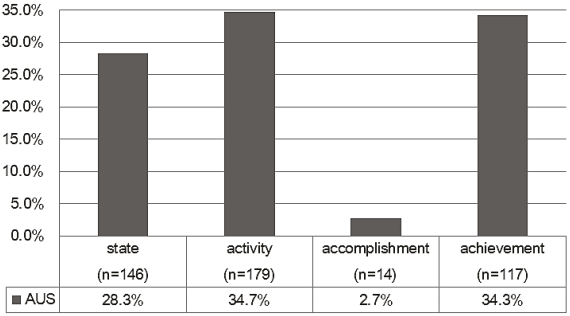


Figure 5.3.5. Aktionsart categories of main verbs in PrPf constructions: ICE-AUS

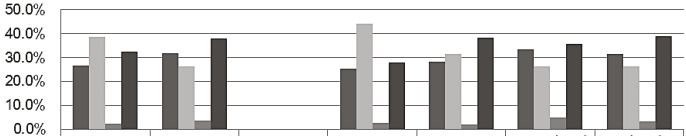
The analysis shows that activity and achievement verbs, as in (5) and (6) respectively, are the most frequent option, followed by state verbs, as exemplified in (7), which account for approximately three out of ten tokens. Accomplishment verbs, as in (8), are very rare.

(5) By all accounts the young Bungendore was a rather turbulent place but it's a town which has *aged* gracefully (ICE-AUS s2b-031)

(6) My wife and I have been using the plant for over three years and have not *noticed* any side effects, in fact it has been beneficial in a number of other ways. (ICE-AUS w2b-026)

(7) Form leading into this race for Fraar has *been* very good (ICE-AUS s2a-017)

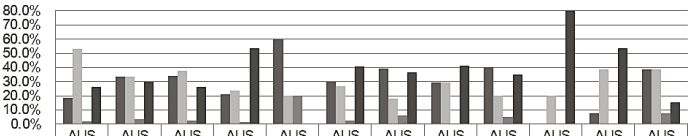
(8) That is the popular view but if you give a person a job and the job has certain prescriptions and it's a job of historic proportions then everybody should have the common decency to understand this understand the demands it makes and make a proper assessment of whether or not the person *has fulfilled* it (ICE-AUS s2a-061)



	SPOKEN (n=356)	WRITTEN (n=160)		dialogues (n=198)	monologues (n=159)	non-printed writing (n=42)	printed writing (n=118)
state	26.7%	31.9%		25.4%	28.3%	33.3%	31.4%
activity	38.5%	26.3%		44.2%	31.4%	26.2%	26.3%
accomplishment	2.2%	3.8%		2.5%	1.9%	4.8%	3.4%
achievement	32.6%	38.1%		27.9%	38.4%	35.7%	39.0%

Figure 5.3.6. Aktionsart categories of main verbs in PrPf constructions: ICE-AUS (register and macro-genre differences)

The register and macro-genre view, as presented in Figure 5.3.6, reveals that the values across the four genres are relatively homogeneous and that none of the distributions differs significantly from the overall figures. However, the spoken macro-genres are less uniform when compared to each other, with dialogues having a markedly higher number of activity verbs. This results in differences approaching the 5% level of statistical significance ($\chi^2 = 7.0908$, df = 3, p = 0.069, $\varphi_C = 0.141$). Speech and writing considered globally also differ nearly significantly ($\chi^2 = 7.7196$, df = 3, p = 0.052, $\varphi_C = 0.122$). This is mainly due to the reverse distribution of state and activity verbs, but no significant differences are found when taking the overall distribution as a baseline.



	AUS S1A (n=107)	AUS S1B (n=90)	AUS S2A (n=88)	AUS S2B (n=71)	AUS W1A (n=5)	AUS W1B (n=37)	AUS W2A (n=33)	AUS W2B (n=34)	AUS W2C (n=20)	AUS W2D (n=5)	AUS W2E (n=13)	AUS W2F (n=13)
state	18.7%	33.3%	34.1%	21.1%	60.0%	29.7%	39.4%	29.4%	40.0%	0.0%	7.7%	38.5%
activity	53.3%	33.3%	37.5%	23.9%	20.0%	27.0%	18.2%	29.4%	20.0%	20.0%	38.5%	38.5%
accomplishment	1.9%	3.3%	2.3%	1.4%	20.0%	2.7%	6.1%	0.0%	5.0%	0.0%	0.0%	7.7%
achievement	26.2%	30.0%	26.1%	53.5%	0.0%	40.5%	36.4%	41.2%	35.0%	80.0%	53.8%	15.4%

Figure 5.3.7. Aktionsart categories of main verbs in PrPf constructions: ICE-AUS (detailed text categories)

Not surprisingly, the detailed representation for the ICE text categories in Figure 5.3.7 shows considerable variability across the various types. Worthy of note is that activity Aktionsart, the category that represents the most verbs in terms of relative frequencies, is in fact the majority variant in a restricted set of (mostly spoken) text types (s1a, s1b, s2a, w2f) only. Overall, activity verbs nevertheless constitute the majority variant, which is due to the fact that the spoken sections of ICE are bigger than the written ones by default. The least frequent category, accomplishment verbs, is absent from a range of text types (popular (w2b), instructional (w2d), and persuasive (w2e) writing), although sample sizes are not necessary small in these three informational written categories (e.g. 34 tokens in w2b). The same applies to achievement verbs in student writing (w1a). Low token counts, however, may have led to some odd results, for instance in respect of the high proportion of achievement Aktionsart in instructional writing (w2d; 5 tokens) or the idiosyncratic distribution of categories in student writing (w1a; 5 tokens); thus, the data for these are merely suggestive. Still, the analysis of the ICE-AUS dataset reveals that text type clearly influences the distributional values of Aktionsart categories in AusE as well.

5.3.3 Sentence type

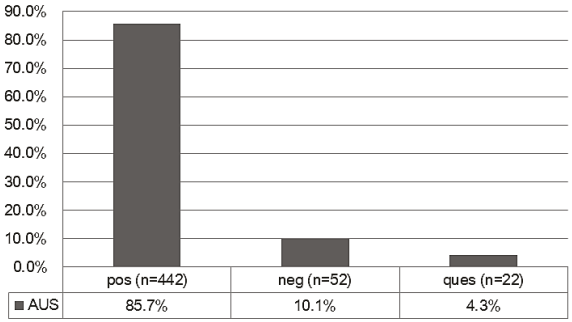


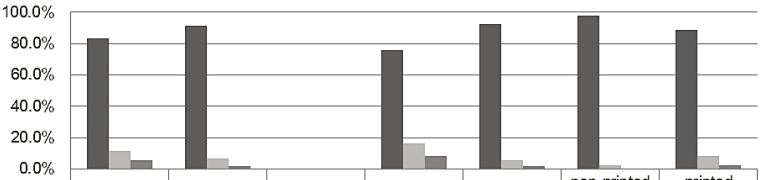
Figure 5.3.8. Distribution of sentence types: ICE-AUS

With regard to the distribution of sentence types in the ICE-AUS sample, positives again emerge as a highly frequent category, covering almost nine out of ten PrPf tokens (see Figure 5.3.8). Every tenth token

occurs in negative statements, as exemplified in (9), and questions with the PrPf, as in (10), occur very rarely.

(9) I haven't been in touch with them since last week end. (ICE-AUS w1b-004)

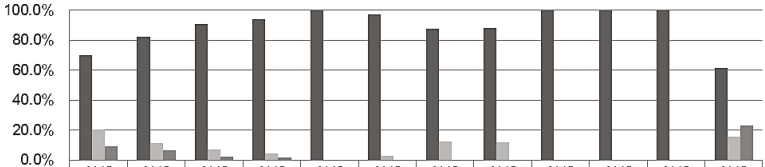
(10) Have you read this article on frozen embryos (ICE-AUS s1a-056)



	SPOKEN (n=356)	WRITTEN (n=160)		dialogues (n=198)	monologues (n=159)	non-printed writing (n=42)	printed writing (n=118)
pos	83.1%	91.3%		75.6%	92.5%	97.6%	89.0%
neg	11.5%	6.9%		16.2%	5.7%	2.4%	8.5%
ques	5.3%	1.9%		8.1%	1.9%	0.0%	2.5%

Figure 5.3.9. Distribution of sentence types: ICE-AUS (register and macro-genre differences)

As expected, positives represent the majority option across both registers and also across all macro-genres, as Figure 5.3.9 shows. Speech and writing differ significantly ($\chi^2 = 6.3098$, df = 2, $p < 0.05$, $\varphi_c = 0.111$) in this variety due to the higher proportions in negatives and questions in the former mode of discourse. To be more precise, this is mainly due to the distribution of the sentence types in dialogues. This interactive genre is conspicuous as the proportions of both negatives and questions are comparatively high and differ significantly ($\chi^2 = 17.959$, df = 2, $p < 0.001$, $\varphi_c = 0.225$) from the distribution for monologues, the other spoken genre, while the printed macro-genres are internally homogeneous when contrasted with each other. At the same time, dialogues is the only macro-genre that differs significantly from the overall distribution ($\chi^2 = 10.3116$, df = 2, $p < 0.01$, $\varphi_c = 0.12$). Note further that questions are completely absent from non-printed writing.



	AUS S1A (n=107)	AUS S1B (n=90)	AUS S2A (n=88)	AUS S2B (n=71)	AUS W1A (n=5)	AUS W1B (n=37)	AUS W2A (n=33)	AUS W2B (n=34)	AUS W2C (n=20)	AUS W2D (n=5)	AUS W2E (n=13)	AUS W2F (n=13)
pos	70.1%	82.2%	90.9%	94.4%	100.0%	97.3%	87.9%	88.2%	100.0%	100.0%	100.0%	61.5%
neg	20.6%	11.1%	6.8%	4.2%	0.0%	2.7%	12.1%	11.8%	0.0%	0.0%	0.0%	15.4%
ques	9.3%	6.7%	2.3%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.1%

Figure 5.3.10. Distribution of sentence types: ICE-AUS (detailed text categories)

A look at the detailed ICE text categories, as presented in Figure 5.3.10, reveals that creative writing (w2f) is the only written text type where questions occur. The high proportion is due to the relatively small size of the sample for this category (13 occurrences of the PrPf, 3 questions) and the fact that the three instances occur where dialogue is simulated (see examples (11) and (12) below).

(11) 'What've I done?' Billy wailed. 'Where're y' takin' me?' 'Down to the station, Billy, we'll sort it all out down there.' (ICE-AUS w2f-004)

(12) "You blew twenty thousand dollars?" "Well, not just me. Friends, you know. I threw a few parties. And I guess the word got round because last week –" "It's all gone?" Cass is awestruck. "That entire amount?" "'Fraid so." (ICE-AUS w2f-017)[249]

Negatives are absent from the written categories reportage (w2c), student (w1a), instructional (w2d) and persuasive (w2e) writing, but highly frequent in the dialogic text types (s1a and s1b) and moreover in the written categories academic (w2a), popular (w2b) and most markedly in creative (w2f) writing. This does not come as a surprise as the latter text type arguably simulates dialogue, the genre with the highest proportion of negatives, in the present case.

[249] Arguably, the clitic in *It's all gone?* could also be interpreted as *is*.

All in all, the distribution of sentence types where PrPfs occur in AusE is clearly sensitive to mode of discourse and further varies when applying a more fine-grained categorization into macro-genres and individual text types.

5.3.4 Semantics

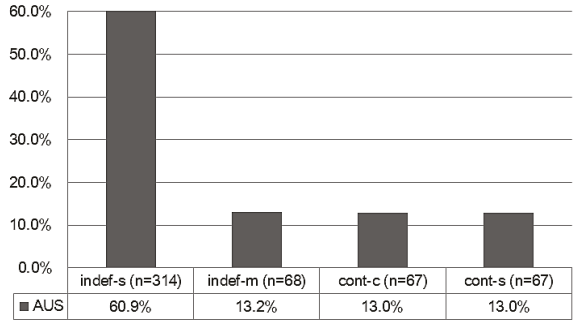
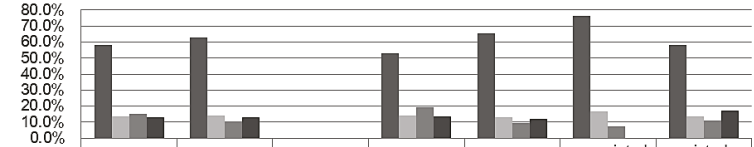


Figure 5.3.11. Distribution of semantic categories: ICE-AUS

As regards the distribution of semantic categories, Figure 5.3.11 shows that in the ICE-AUS data, too, the indefinite single act/event reading (indef-s), as in (13), constitutes the major variant, with three out of five occurrences. The remaining categories, indefinite multiple acts/events (indef-m), as exemplified in (14), continuative acts/events (cont-c), as in (15) and continuative states (cont-s), as in (16), are distributed evenly.

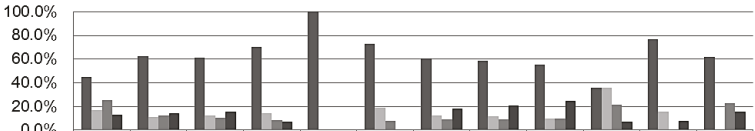
(13) The threat to properties at Fox Valley has eased (ICE-AUS s2b-020)
(14) History has made artists of all these painters who lack energy and feature. (ICE-AUS w2a-004)
(15) Most people to date have taken two leaves daily but as a result of additional information which I now have, my wife and I have commenced taking three leaves, about the size of a 20 cent coin, a day. (ICE-AUS w2b-026)
(16) Yep that's always been that's been his platform (ICE-AUS s1a-001)



	SPOKEN (n=356)	WRITTEN (n=160)		dialogues (n=198)	monologues (n=159)	non-printed writing (n=42)	printed writing (n=118)
indef-s	58.4%	62.7%		52.8%	65.4%	76.2%	58.3%
indef-m	13.8%	14.2%		14.2%	13.2%	16.7%	13.4%
cont-c	14.9%	10.1%		19.3%	9.4%	7.1%	11.0%
cont-s	12.9%	13.0%		13.7%	11.9%	0.0%	17.3%

Figure 5.3.12. Distribution of semantic categories: ICE-AUS (register and macro-genre differences)

It is obvious from Figure 5.3.12 that the spoken and written sections of the Australian corpus component overall are relatively homogeneous. However, the distribution of the semantic readings of the two spoken macro-genres yields significant differences when compared to each other ($\chi^2 = 8.4121$, df = 3, $p < 0.05$, $\varphi_c = 0.154$). Likewise, the two written macro-genres are also significantly different from each other ($\chi^2 = 9.6038$, df = 3, $p < 0.05$, $\varphi_c = 0.238$), as is non-printed writing from the overall distribution ($\chi^2 = 8.5178$, df = 3, $p < 0.05$, $\varphi_c = 0.123$). These findings clearly indicate that distributions of semantic readings of the PrPf in ICE-AUS are susceptible to genre effects.



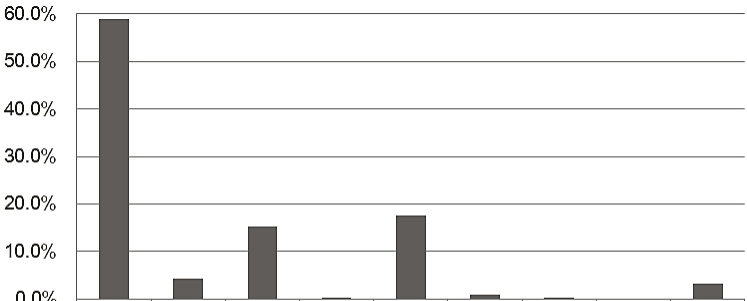
	AUS S1A (n=107)	AUS S1B (n=90)	AUS S2A (n=88)	AUS S2B (n=71)	AUS W1A (n=5)	AUS W1B (n=37)	AUS W2A (n=33)	AUS W2B (n=34)	AUS W2C (n=20)	AUS W2D (n=5)	AUS W2E (n=13)	AUS W2F (n=13)
indef-s	44.9%	62.2%	61.4%	70.4%	100.0%	73.0%	60.6%	58.8%	55.0%	35.7%	76.9%	61.5%
indef-m	16.8%	11.1%	12.5%	14.1%	0.0%	18.9%	12.1%	11.8%	10.0%	35.7%	15.4%	0.0%
cont-c	25.2%	12.2%	10.2%	8.5%	0.0%	8.1%	9.1%	8.8%	10.0%	21.4%	0.0%	23.1%
cont-s	13.1%	14.4%	15.9%	7.0%	0.0%	0.0%	18.2%	20.6%	25.0%	7.1%	7.7%	15.4%

Figure 5.3.13. Distribution of semantic categories: ICE-AUS (detailed text categories)

Figure 5.3.13 illustrates that the picture becomes even more variable when individual text types are considered. While indef-s readings are the

majority variant across all text types and the exclusive occurrence of indef-s in student writing (w1a) can be ascribed to the low number of tokens (5) for this category, variability is more pronounced within the written text categories. There we can find the most extreme – that is, both the highest and lowest relative values – for all four semantic categories (apart from the highest value for cont-c, which occurs in private dialogues (s1a)) and this impression is further confirmed by the SD across all respective categories, which is on average twice as high ($\sigma = 0.11$) for the written categories in contrast to the value for the spoken ones ($\sigma = 0.05$).

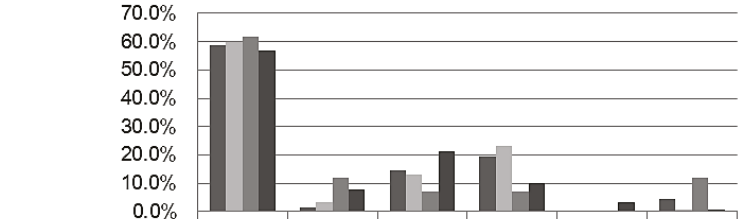
5.3.5 Preceding time reference



	PrTs (n=304)	PrTp (n=22)	SPst (n=78)	Pstp (n=1)	SPrPf (n=90)	PrPfp (n=4)	PaPfs (n=1)	PaPfp (n=0)	F (n=16)
AUS	58.9%	4.3%	15.1%	0.2%	17.4%	0.8%	0.2%	0.0%	3.1%

Figure 5.3.14. TR forms preceding the PrPf: ICE-AUS

The distribution of preceding TR forms in the AusE data is displayed in Figure 5.3.14. Almost three fifths of all PrPfs follow a PrTs form, while approximately one sixth of all tokens are preceded by SPrPf and one out of seven by SPst. Other preceding TR forms, apart from PrTp and F, are negligible.



	PrTs (n=304)	PrTp (n=22)	SPst (n=78)	SPrPf (n=90)	PrPfp (n=4)	F (n=16)
dialogues	58.9%	1.5%	14.7%	19.3%	0.0%	4.6%
monologues	59.7%	3.1%	13.2%	23.3%	0.0%	0.6%
non-printed writing	61.9%	11.9%	7.1%	7.1%	0.0%	11.9%
printed writing	56.8%	7.6%	21.2%	10.2%	3.4%	0.8%

Figure 5.3.15. TR forms preceding the PrPf: ICE-AUS (macro-genre differences)

Figure 5.3.15, which shows the results for the various macro-genres, establishes a number of further points. First, as could be expected from the overall distribution, the highly frequent forms PrTs, SPst and SPrPf account for the vast majority of all preceding TR (75% in non-printed writing and nine out of ten tokens or more in the remaining genres). Second, considerable variation can be seen in the macro-genres, mainly due to noticeable differences in the proportions of TR forms other than PrTs. While the spoken genres are relatively homogeneous (i.e. not differing significantly) in respect to both the overall distribution and when compared among themselves, both non-printed ($\chi^2 = 17.405$, df = 7, $p < 0.05$, $\varphi_c = 0.177$) and, less obviously, printed ($\chi^2 = 15.111$, df = 7, $p < 0.05$, $\varphi_c = 0.154$) writing differ significantly from the overall distribution. The same holds when the two modes of discourse are directly contrasted ($\chi^2 = 30.611$, df = 7, $p < 0.001$, $\varphi_c = 0.244$; relative values not shown in Figure 5.3.15). Expressed in relative proportions of TR forms, the written genres yield high proportions of PrTp and low proportions of SPrPf, while their values diverge considerably in the percentage of preceding SPst, where the value for printed writing is three times as high as in non-printed writing. Note also the high proportion of F in the latter macro-genre, although this result has to be interpreted with caution due to the small number of tokens (42 in total, 5 preceded by F) from non-

printed sources. In any case, as a whole, the findings from this section suggest strong genre effects, predominantly in the written sections.

5.3.6 Perfect-friendliness and text type effects

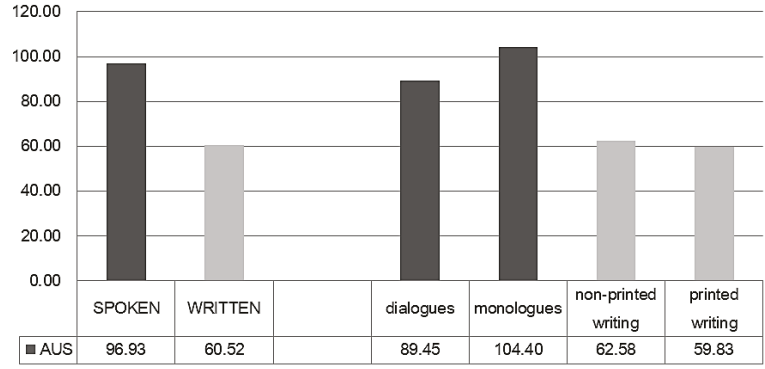


Figure 5.3.16. Frequency of PrPf occurrences in ICE-AUS registers and macro-genres (normalized frequency per 10,000 words)

As revealed in Figure 5.3.16, AusE is broadly speaking more PrPf-friendly (73 PrPfs/pttw) than the average across all varieties (69). It patterns with a number of other varieties in that markedly more PrPfs surface in speech compared to writing and that, within the former category, the macro-genre of monologues shows the highest PrPf frequencies. The two written genres yield similar values. It has been proposed above that the definite temporal adverbial constraint is suspended in AusE and that the PrPf therefore may possess an extended functional range. Whether this is actually the case is further investigated below (see Section 6.6).

The Australian data are dispersed around the overall average value (σ = 31.90), with the textual categories instructional (w2d; 26) and creative (w2f; 33) writing representing low frequency types that contrast with scripted monologues (s2b; 121) and letters (w1b; 107) at the other end of the spectrum. Surprisingly, IndE represents the variety that correlates most strongly with AusE as regards the normalized frequencies for the individual text categories (r = 0.84), while other strong correlations

appear with IrE (r = 0.78) and BrE (r = 0.77), both varieties that AusE has strongly been related to from a historical and linguistic point of view.

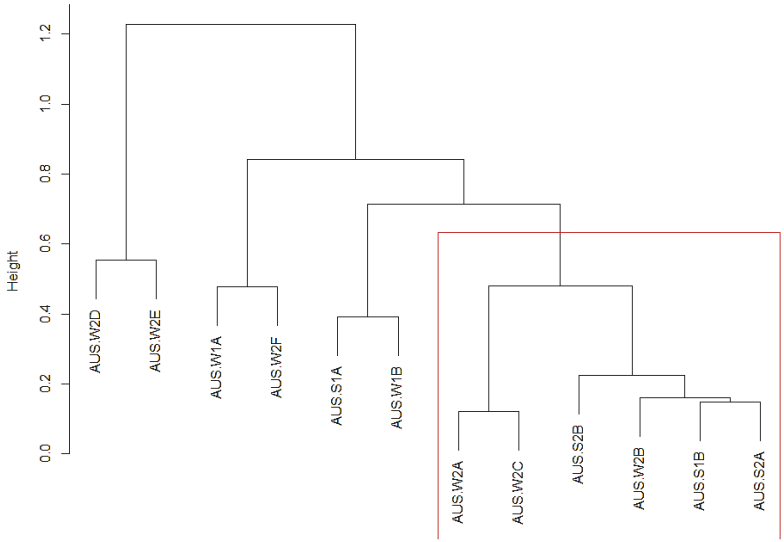


Figure 5.3.17. Cluster dendrogram of similarity across ICE text categories in AusE

The cluster dendrogram (c = 0.86) in Figure 5.3.17 shows that three out of the four spoken categories cluster together closely toward the right of the representation in a significant cluster (highlighted in the Figure). The remaining spoken text type, private dialogues (s1a), forms a cluster with letters (w1b). A possible explanation for the latter finding is that social letters in particular resemble private dialogues in their topics and thus arguably also in their informal style (cf. examples (17) and (18)), which may have repercussions on the usage of TR forms. Still, the cluster is not overly distant from that containing the other spoken categories.

(17) Am so sorry I haven't written more, its It's hard to realise how little I've written to you - it has been incredibly hectic here, as usual, we seem to have to constantly make decisions to say NO to people, which is hard: Our biggest 'NO' was

sending Taj back to the valleys after his final term exams. (ICE-AUS w1b-002)

(18) I've had the wog you were talking about + I think it was short circuited by me taking anti biotics for mey arm. (ICE-AUS w1b-003)

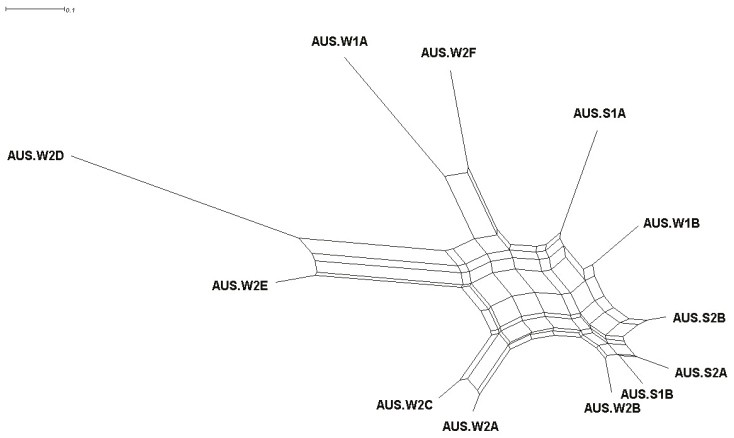


Figure 5.3.18. NeighborNet of similarity across ICE text categories in AusE

The NeighborNet view, as displayed in Figure 5.3.18, provides further insights into the similarity of text categories as far as PrPf usage is concerned. In this mode of representation, all of the spoken categories apart from private dialogues (s1a) can be found on neighboring nodes, while letters (w1b), the involved text type most likely to resemble speech, is also not distant. This group (spoken texts + w1b) shares an extended number of splits. Two pairs of written text categories, student (w1a) and creative (w2f) writing as well as persuasive (w2e) and, most markedly, instructional (w2d) writing, are distant from the remaining text types. Therefore, we can again conclude that Biber & Conrad's (2009: 261) hypothesis that spoken texts are more similar to each other than written texts is applicable in terms of the distribution of factors determining PrPf usage in AusE.

5.4 New Zealand English (phase 5)

The situation in New Zealand can be related to that in Australia as regards both the settlement history of this former colony and many of its linguistic features. However, a number of differences exist. Above all, New Zealand was not used as a penal colony and was scarcely populated by white settlers until 1840, when the country was formally established as a Crown colony in the Treaty of Waitaingi, agreed between the settlers and the indigenous Maoris. Political independence from Britain was gained stepwise at the beginning of the 20th century, when New Zealand was granted dominion status, and eventually in 1947, when the country became independent. Some observers suggest that from the 1970s onwards, the economic orientation has shifted away from Europe to the Pacific region and Asia (Schneider 2007: 131).

Commonly, southern English dialects besides Cockney, Irish and Scottish varieties of English are seen as the main input for NZE (Hickey 2004: 18; Schreier 2010: 460), and a later reluctance to accept AmE influences with a corresponding adherence to BrE norms that has been more stable than in AusE, the Antipodean sister variety, has been reported for NZE (Köppl 1983a: 148). In contrast, others have suggested a marked influence of AmE (and also of CanE; Leitner 1992: 207) due to strong economic and cultural ties (Allerton et al. 2002: XII; Hundt 2002: 63) since the beginning of the 20th century, in particular as regards lexical items (Hay et al. 2008: 75–76). In addition, mutual influence with AusE is likely (Hundt 2002: 70–71). However, NZE now is conceived as an independent variety. It is distinct from others mainly in the areas of phonology and lexis and should be seen on a par with other "national" L1 varieties such as AmE or BrE (Hundt et al. 2004: 560), even if the evidence for an influence of NZE on neighboring varieties of English is weak (Hundt & Biewer 2007: 263). Equally, NZE is becoming more distinct from AusE (Schneider 2007: 132). With regard to dialectal diversification, NZE is said to be lagging behind AusE, so that present-day NZE is still relatively uniform (Schreier 2010: 460), although regional differences are emerging, again mainly in the area of phonology (Schneider 2007: 133). It remains to be tested whether NZE also shows some distinctive traits in the area of morphosyntax under investigation.

Like in AusE, in NZE, too, there are hints of an increased use of the PrPf in contexts where in standard varieties the SPst would be used (Quinn 1999: 196), while opinions are divided whether the feature is to be categorized as non-standard and restricted to speech (Hundt 1998: 74; Hundt et al. 2004: 568) or also accepted in writing and in more formal contexts (Hay et al. 2008: 51). Either way, this generalized usage of the PrPf suggests that NZE should be a relatively PrPf-friendly variety, at least in speech and in informal contexts.

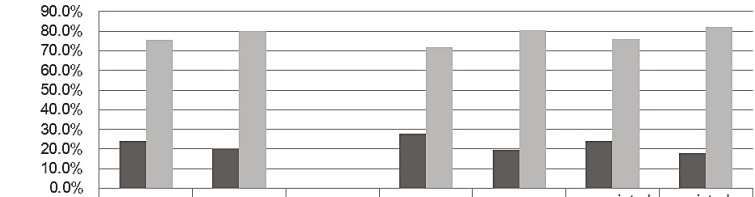
A methodological note of caution is in order due to another salient feature of spoken NZE in the area of PrPf marking. It has been noted that speakers of NZE may occasionally drop auxiliary *have* (Hundt et al. 2004: 583–584; Hay et al. 2008: 50) with high-frequency verbs, giving rise to lone past participles as in (1) or (2), which in some cases may be congruent with the corresponding SPst form.

(1) we never *seen* where she lived (ICE-NZ s1a-018)
(2) you *been* right to the top or cape reinga and that sort of thing (ICE-NZ s1a-067)[250]

These reduced forms, which are rarely produced by educated speakers and consequently also occur rarely in the ICE data, were usually identified as SPst forms by the POS-tagger and are therefore excluded from the ensuing analysis of the PrPf (but cf. Section 6.4). Alternatively, this could be interpreted as leveling between SPst and past participle forms in which the latter replaces the former (eWAVE Feature 131 2012).

[250] Due to the transcription conventions for ICE-NZ, no capitalization is used in the spoken parts of this corpus component.

5.4.1 Temporal adverbials

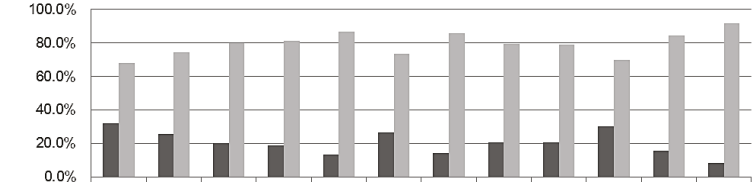


	SPOKEN (n=334)	WRITTEN (n=235)		dialogues (n=186)	monologues (n=148)	non-printed writing (n=83)	printed writing (n=152)
with TA	24.3%	20.0%		28.0%	19.6%	24.1%	17.8%
without TA	75.7%	80.0%		72.0%	80.4%	75.9%	82.2%

Figure 5.4.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-NZ (register and macro-genre differences)

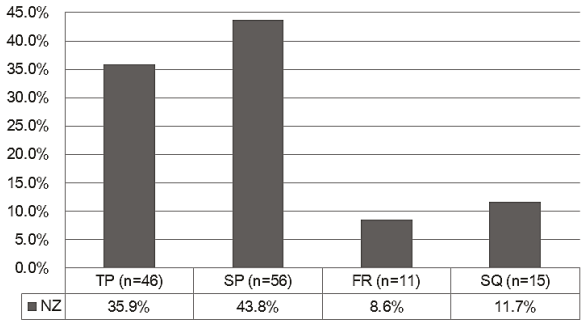
The overall proportion of PrPf occurrences that is temporally specified in the ICE-NZ data amounts to more than every fifth instance and the breakdown for the registers and the macro-genres is shown in Figure 5.4.1. It is evident that temporal specification is more likely in speech, while the two modes of discourse do not differ significantly from each other. Further, none of the macro-genres is significantly different from the overall distribution. However, note that while the written genres are internally homogeneous, the proportion of temporally specified instances in dialogues, where almost three out of ten PrPfs are specified by a temporal adverbial, clearly exceeds the value for monologues, although differences are not significant ($\chi^2 = 3.1375$, df = 1, p = 0.077).

The proportions of the individual text types are displayed in Figure 5.4.2. As could be expected from the results of a number of other varieties, private dialogues (s1a) have the highest rate of temporal specification, while variability in the proportional values of specified tokens is more pronounced within the informational text categories, ranging from 8% in creative writing (w2f) to 30% in instructional writing (w2d). Thus, it can be hypothesized that the distributions are robustly determined by text type.



	NZ S1A (n=72)	NZ S1B (n=114)	NZ S2A (n=84)	NZ S2B (n=64)	NZ W1A (n=15)	NZ W1B (n=68)	NZ W2A (n=42)	NZ W2B (n=29)	NZ W2C (n=24)	NZ W2D (n=20)	NZ W2E (n=13)	NZ W2F (n=24)
with TA	31.9%	25.4%	20.2%	18.8%	13.3%	26.5%	14.3%	20.7%	20.8%	30.0%	15.4%	8.3%
without TA	68.1%	74.6%	79.8%	81.3%	86.7%	73.5%	85.7%	79.3%	79.2%	70.0%	84.6%	91.7%

Figure 5.4.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-NZ (detailed text categories)



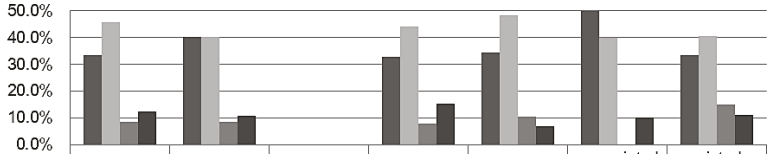
	TP (n=46)	SP (n=56)	FR (n=11)	SQ (n=15)
NZ	35.9%	43.8%	8.6%	11.7%

Figure 5.4.3. Types of PrPf adverbial specification: ICE-NZ

Figure 5.4.3 reveals that adverbials of span and duration (SP), as exemplified in (3), represent the most common adverbial category, with more than two fifths of all specified tokens. Adverbials of time-position (TP), as in (4), still cover more than one third of all relevant occurrences, while the proportion of adverbials of sequence (SQ), as in (5), exceeds one tenth. Frequency adverbials (FR), as in (6), are the least common category in the ICE-NZ sample.

(3) Pile driving technology has advanced *since that time*. (ICE-NZ w2a-034)
(4) i've talked to her *this afternoon* (ICE-NZ s1b-074)
(5) haven't you burnt it *yet* (ICE-NZ s1b-074)

(6) Mr Moore has said *repeatedly* he will not launch a leadership bid against Mr Palmer. (ICE-NZ w2e-007)



	SPOKEN (n=81)	WRITTEN (n=47)		dialogues (n=52)	monologues (n=29)	non-printed writing (n=20)	printed writing (n=27)
TP	33.3%	40.4%		32.7%	34.5%	50.0%	33.3%
SP	45.7%	40.4%		44.2%	48.3%	40.0%	40.7%
FR	8.6%	8.5%		7.7%	10.3%	0.0%	14.8%
SQ	12.3%	10.6%		15.4%	6.9%	10.0%	11.1%

Figure 5.4.4. Types of PrPf adverbial specification: ICE-NZ (register and macro-genre differences)

As can be seen in Figure 5.4.4, writing shows a larger proportion of TP adverbials. However, the distribution is not significantly different from that found in speech. The most frequent type of temporal specification is by SP adverbials in all macro-genres apart from non-printed writing. The latter genre is also odd in that no FR adverbials appear. Still, although the relative values suggest the contrary, the distribution of temporal specification categories in this genre is not significantly different from the overall values.[251]

On the basis of the NZE data from our ICE sample, the hierarchy of temporal adverbial categories that typically occur when PrPfs are temporally specified goes along the lines SP > TP > SQ > FR. This finding indicates that SP and TP can be understood as adverbial types that characteristically (nearly three quarters of all cases) surface with PrPfs in NZE.

[251] The idiosyncratic behavior in the relative values as shown in the figure are most likely due to the fact that non-printed writing in the ICE-NZ sample is the macro-genre with the lowest number of tokens (20).

5.4.2 Aktionsart

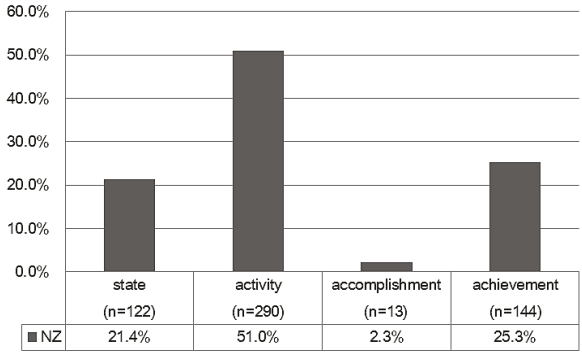


Figure 5.4.5. Aktionsart categories of main verbs in PrPf constructions: ICE-NZ

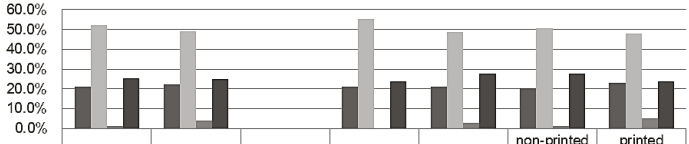
Figure 5.4.5 presents the overall distribution of Aktionsart categories in ICE-NZ. In this dataset, more than half of all tokens belong to the activity category, as shown in example (7). Approximately one in four instances of the PrPf has an achievement Aktionsart, as in (8), and state verbs, as in (9), cover more than one in five tokens. Accomplishment verbs, as in example (10), once again represent the least frequent category.

(7) Dr Sutton has *worked* hard on a drought relief package, but the benefits of that, all things being equal and they seldom are, will not be felt till July-August. (ICE-AUS w2c-006)

(8) well we should after we've *finished* we should go in sit in his room (ICE-NZ s1a-039)

(9) i've *forgotten* now it was bloody hard though (ICE-NZ s1a-036)

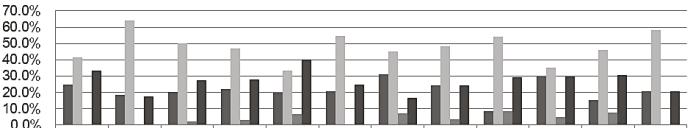
(10) [...] a number who're connected with um television channels production companies or whatever *have become* s somewhat defensive about what they've er they've done (ICE-NZ s2b-045)



	SPOKEN (n=334)	WRITTEN (n=235)		dialogues (n=186)	monologues (n=148)	non-printed writing (n=83)	printed writing (n=152)
state	21.0%	22.1%		21.0%	20.9%	20.5%	23.0%
activity	52.4%	48.9%		55.4%	48.6%	50.6%	48.0%
accomplishment	1.2%	3.8%		0.0%	2.7%	1.2%	5.3%
achievement	25.4%	25.1%		23.7%	27.7%	27.7%	23.7%

Figure 5.4.6. Aktionsart categories of main verbs in PrPf constructions: ICE-NZ (register and macro-genre differences)

What emerges from Figure 5.4.6, which shows the breakdown of Aktionsart categories for the two registers and the four macro-genres, is a homogeneous picture. The only genre that is conspicuous is dialogues, with the highest relative value for activities but no accomplishments. Yet differences are not significant compared either to the overall or speech distribution or when comparing the two spoken genres.



	NZ S1A (n=72)	NZ S1B (n=114)	NZ S2A (n=84)	NZ S2B (n=64)	NZ W1A (n=15)	NZ W1B (n=68)	NZ W2A (n=42)	NZ W2B (n=29)	NZ W2C (n=24)	NZ W2D (n=20)	NZ W2E (n=13)	NZ W2F (n=24)
state	25.0%	18.4%	20.2%	21.9%	20.0%	20.6%	31.0%	24.1%	8.3%	30.0%	15.4%	20.8%
activity	41.7%	64.0%	50.0%	46.9%	33.3%	54.4%	45.2%	48.3%	54.2%	35.0%	46.2%	58.3%
accomplishment	0.0%	0.0%	2.4%	3.1%	6.7%	0.0%	7.1%	3.4%	8.3%	5.0%	7.7%	0.0%
achievement	33.3%	17.5%	27.4%	28.1%	40.0%	25.0%	16.7%	24.1%	29.2%	30.0%	30.8%	20.8%

Figure 5.4.7. Aktionsart categories of main verbs in PrPf constructions: ICE-NZ (detailed text categories)

The more detailed view, which presents the distributions for the individual ICE text types (see Figure 5.4.7), is also relatively homogeneous, although naturally more variability is in evidence. Student writing (w1a) is the only text type where activity verbs are not the most frequent option, while the rare category of accomplishment verbs is absent from four of the text types (private and public dialogues (s1a, s1b), letters

(w1b) and creative writing (w2f)). Another result worth reporting is the comparatively low proportion of state verbs in reportage (w2c). As sample sizes for the individual categories are 15 tokens or more, we can conclude that text type decidedly influences the distributions of the Aktionsart categories in the ICE-NZ data.

5.4.3 Sentence type

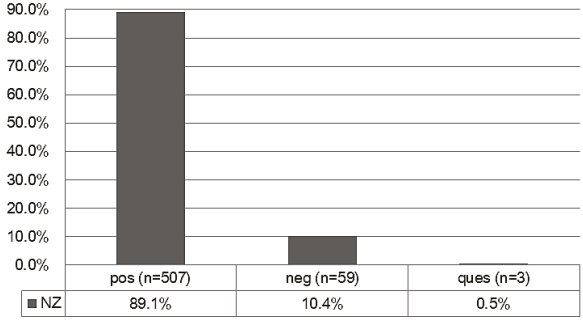
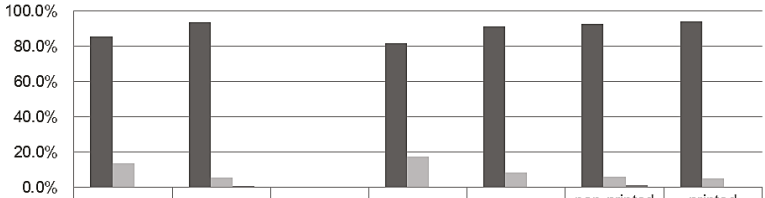


Figure 5.4.8. Distribution of sentence types: ICE-NZ

The distribution of sentence types in NZE, which is shown in Figure 5.4.8, is as could be expected from the analysis of other varieties. Positives cover nearly nine out of ten PrPf occurrences, while the remaining instances, apart from a few questions (3 tokens in total, see example (11), which is even of a rhetorical nature), can be categorized as negatives (see example (12)).

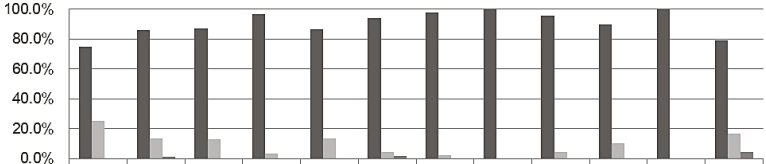
(11) I suppose you've spent all weekend on the bowling green? (ICE-NZ w1b-008)
(12) Sue has been deep in university study and has not had the time for bridge [...] (ICE-NZ w2d-015)



	SPOKEN (n=334)	WRITTEN (n=235)		dialogues (n=186)	monologues (n=148)	non-printed writing (n=83)	printed writing (n=152)
pos	85.9%	93.6%		81.7%	91.2%	92.8%	94.1%
neg	13.8%	5.5%		17.7%	8.8%	6.0%	5.3%
ques	0.3%	0.9%		0.5%	0.0%	1.2%	0.7%

Figure 5.4.9. Distribution of sentence types: ICE-NZ (register and macro-genre differences)

Figure 5.4.9 shows that positives again are by far the most frequent sentence type in both modes of discourse and across all genres. Due to the fact that negatives are nearly three times as frequent in speech, this register differs significantly from writing here ($\chi^2 = 10.745$, df = 2, $p < 0.005$, $\varphi_c = 0.137$). As in AusE, this is caused by the large proportion of negatives in dialogues, which, likewise, also renders dialogues significantly different from monologues within the spoken sections ($\chi^2 = 6.4629$, df = 2, $p < 0.05$, $\varphi_c = 0.139$). Yet another parallel to the findings for the Antipodean sister variety is that the written macro-genres are internally homogeneous.



	NZ S1A (n=72)	NZ S1B (n=114)	NZ S2A (n=84)	NZ S2B (n=64)	NZ W1A (n=15)	NZ W1B (n=68)	NZ W2A (n=42)	NZ W2B (n=29)	NZ W2C (n=24)	NZ W2D (n=20)	NZ W2E (n=13)	NZ W2F (n=24)
pos	75.0%	86.0%	86.9%	96.9%	86.7%	94.1%	97.6%	100.0%	95.8%	90.0%	100.0%	79.2%
neg	25.0%	13.2%	13.1%	3.1%	13.3%	4.4%	2.4%	0.0%	4.2%	10.0%	0.0%	16.7%
ques	0.0%	0.9%	0.0%	0.0%	0.0%	1.5%	0.0%	0.0%	0.0%	0.0%	0.0%	4.2%

Figure 5.4.10. Distribution of sentence types: ICE-NZ (detailed text categories)

Figure 5.4.10 presents the proportions of sentence types for the individual text categories. It indicates, first of all, that the rare category of questions appears in three text categories only and is apparently most salient in creative writing (w2f). This finding has to be treated with caution, however, as relative proportions due to varying absolute sample sizes may be misleading here (merely 1 token in each relevant text type occurs as a question). Furthermore, no negatives could be retrieved in popular (w2b) and persuasive writing (w2e), but are much more frequent in the more involved categories dialogues (s1a, s1b), unscripted monologues (s2a) and student writing (w1a) besides creative writing (w2f), which may imitate dialogue (see example (13) below), as was also shown for AusE.

(13) 'What about his mother?' asked Mrs Yamada. 'I haven't met her, but my aunt has and says she's very nice. [...]' (ICE-NZ w2f-010)

In sum, the data suggest that the distribution of sentence types in PrPf contexts is above all determined by mode of discourse. In addition, the more detailed analysis showed evidence of variability according to macro-genres and, moreover, individual textual categories in ICE-NZ.

5.4.4 Semantics

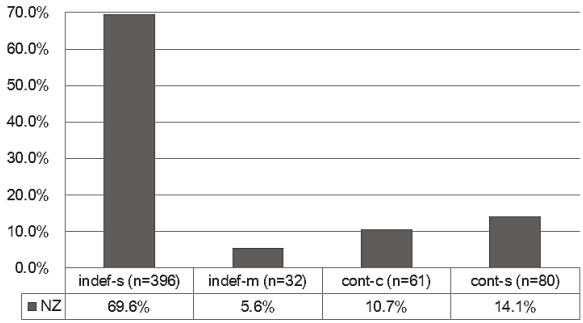
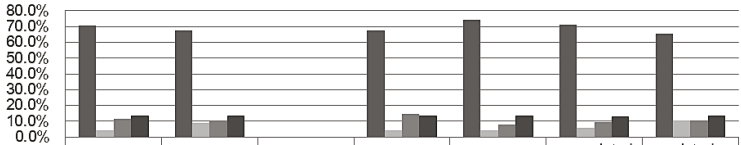


Figure 5.4.11. Distribution of semantic categories: ICE-NZ

Figure 5.4.11 displays the breakdown of the semantic categories for NZE. It emerges that indefinite single acts/events (indef-s), which are exemplified by (14), with nearly seven out of ten occurrences are again the most frequent variant. Indefinite multiple act/event readings (indef-m), as in (15), are the least common category in this variety, while continuative acts/events (cont-c), as in example (7) (for convenience repeated here as (16)), and continuative states (cont-s), as in (17), also constitute minor options.

(14) and er eroni clarke has pulled off the tackle of the match so far (ICE-NZ s2a-016)
(15) we've lost a job every ten minutes (ICE-NZ s1b-032)
(16) Dr Sutton has worked hard on a drought relief package, but the benefits of that, all things being equal and they seldom are, will not be felt till July-August. (ICE-NZ w2c-006)
(17) [...] yet some of those factors have been with us some time. (ICE-NZ w2e-002)

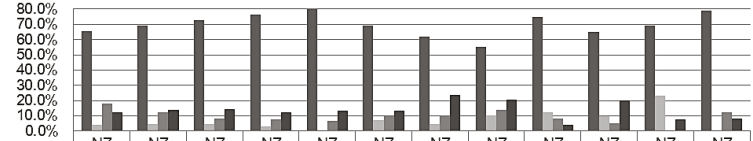


	SPOKEN (n=334)	WRITTEN (n=235)		dialogues (n=186)	monologues (n=148)	non-printed writing (n=83)	printed writing (n=152)
indef-s	70.7%	67.5%		67.7%	74.3%	71.1%	65.6%
indef-m	4.2%	8.9%		4.3%	4.1%	6.0%	10.4%
cont-c	11.7%	10.1%		14.5%	8.1%	9.6%	10.4%
cont-s	13.5%	13.5%		13.4%	13.5%	13.3%	13.6%

Figure 5.4.12. Distribution of semantic categories: ICE-NZ (register and macro-genre differences)

Figure 5.4.12 reveals that, by and large, the ICE-NZ data are fairly uniform in the distribution of semantic categories. Although the relative proportion of indef-m is more than double in the written section, distributions for speech and writing do not differ significantly and the same applies both to differences between the individual macro-genres and when the values of the four macro-genres are contrasted with those for

the two modes of discourse. In sum, this suggests that genre and register do not determine the distributions of the semantic categories of the PrPf in NZE.



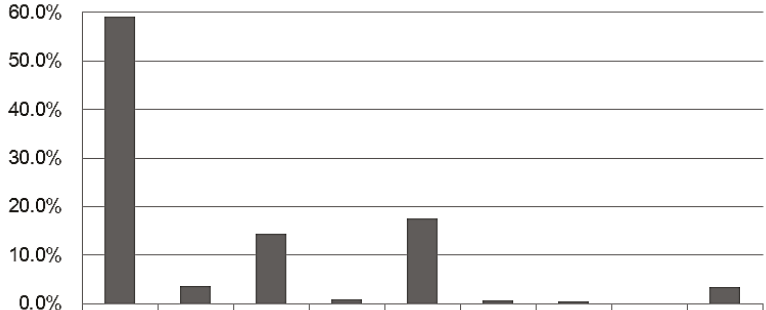
	NZ S1A (n=72)	NZ S1B (n=114)	NZ S2A (n=84)	NZ S2B (n=64)	NZ W1A (n=15)	NZ W1B (n=68)	NZ W2A (n=42)	NZ W2B (n=29)	NZ W2C (n=24)	NZ W2D (n=20)	NZ W2E (n=13)	NZ W2F (n=24)
indef-s	65.3%	69.3%	72.6%	76.6%	80.0%	69.1%	61.9%	55.2%	75.0%	65.0%	69.2%	79.2%
indef-m	4.2%	4.4%	4.8%	3.1%	0.0%	7.4%	4.8%	10.3%	12.5%	10.0%	23.1%	0.0%
cont-c	18.1%	12.3%	8.3%	7.8%	6.7%	10.3%	9.5%	13.8%	8.3%	5.0%	0.0%	12.5%
cont-s	12.5%	14.0%	14.3%	12.5%	13.3%	13.2%	23.8%	20.7%	4.2%	20.0%	7.7%	8.3%

Figure 5.4.13. Distribution of semantic categories: ICE-NZ (detailed text categories)

That the ICE-NZ data are relatively homogeneous can also be seen in the detailed view as presented in Figure 5.4.13. Indefinite single readings are the majority variant across all text types. Slightly less variance is observable in the spoken categories, as is shown by the average SD values, which is more than three times as high for the written ($\sigma = 0.07$) in contrast to the spoken sections ($\sigma = 0.02$). Within texts from the former category, comparatively high proportions of cont-s occur in academic (w2a) and popular (w2b) writing, while indef-m readings show the greatest dispersion ($\sigma = 0.08$), with relative values ranging between zero in student (w1a) and creative (w2f) writing and approximating one fourth of all tokens in instructional (w2d) and persuasive writing (w2e).

5.4.5 Preceding time reference

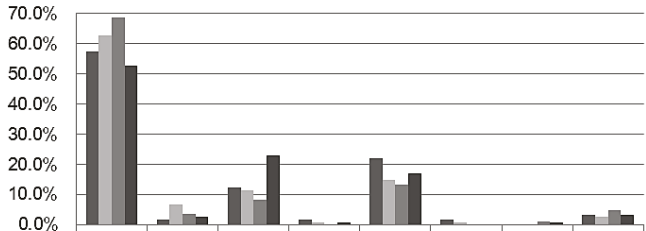
Figure 5.4.14 shows the distribution of preceding TR forms in the ICE-NZ sample. This breakdown is strikingly similar to that found for AusE (see Section 5.3.5). Approximately three out of five tokens are preceded by a PrTs form, every sixth occurrence by SPrPf, and every seventh occurrence by SPst. PrTp and F are the only other TR forms that surface at least to a limited extent.



	PrTs (n=337)	PrTp (n=20)	SPst (n=82)	Pstp (n=5)	SPrPf (n=100)	PrPfp (n=4)	PaPfs (n=2)	PaPfp (n=0)	F (n=19)
NZ	59.2%	3.5%	14.4%	0.9%	17.6%	0.7%	0.4%	0.0%	3.3%

Figure 5.4.14. TR forms preceding the PrPf: ICE-NZ

The analysis of the macro-genres, with relative values displayed in Figure 5.4.15, reveals that preceding TR forms other than PrTs, SPst and SPrPf are rare (highest individual value, PrTp in monologues, < 7%) across all categories.



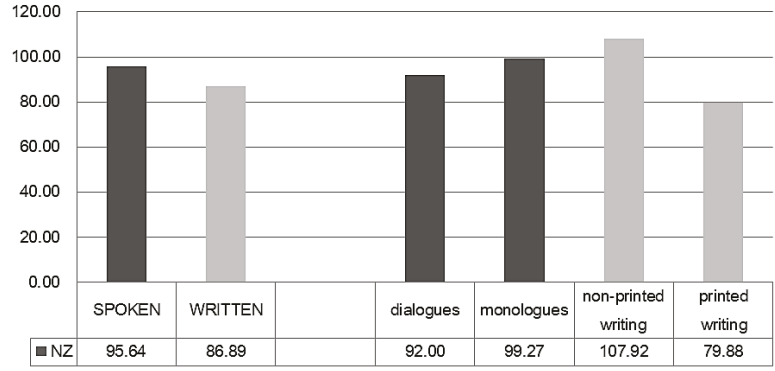
	PrTs (n=337)	PrTp (n=20)	SPst (n=82)	Pstp (n=5)	SPrPf (n=100)	PrPfp (n=4)	PaPfs (n=2)	F (n=19)
dialogues	57.5%	1.6%	12.4%	1.6%	22.0%	1.6%	0.0%	3.2%
monologues	62.8%	6.8%	11.5%	0.7%	14.9%	0.7%	0.0%	2.7%
non-printed writing	68.7%	3.6%	8.4%	0.0%	13.3%	0.0%	1.2%	4.8%
printed writing	52.6%	2.6%	23.0%	0.7%	17.1%	0.0%	0.7%	3.3%

Figure 5.4.15. TR forms preceding the PrPf: ICE-NZ (macro-genre differences)

Again, the ICE-NZ data yield a relatively homogeneous picture with none of the genres differing from the overall distribution or the ones for the two registers (relative values not shown in Figure 5.4.15). Register-internally, however, some minor variances emerge. While the distribu-

tions of the TR forms in the spoken genres are similar (i.e. not significantly different) when contrasted directly with each other, the differences between non-printed and printed writing fall just below the 5% level of statistical significance as determined by a Fisher-exact test ($p < 0.05$), which was used in the present case due to a number of low absolute counts in some cells. Therefore, all in all, we can conclude that effects of register and genre on the distributions of preceding TR forms in the ICE-NZ sample are only very weak at best.

5.4.6 Perfect-friendliness and text type effects



	SPOKEN	WRITTEN		dialogues	monologues	non-printed writing	printed writing
NZ	95.64	86.89		92.00	99.27	107.92	79.88

Figure 5.4.16. Frequency of PrPf occurrences in ICE-NZ registers and macro-genres (normalized frequency per 10,000 words)

As hypothesized above, NZE is indeed a very PrPf-friendly variety (90 PrPfs/pttw) compared to others (average across all varieties = 69). Further, in the ICE-NZ data (see Figure 5.4.16) the PrPf ratio is on average higher in speech than in writing, while non-printed writing is the macro-genre with the highest normalized frequency of PrPfs overall (108). These findings are in line with those of Elsness (2009a: 97) and confirm the above hypothesis that NZE should yield high PrPf-friendliness scores in speech and in informal contexts, though this may stem from reasons other than an extension of the functional range of the PrPf into the domains of the SPst.

The variance of the NZE sample amounts to σ = 25.83 and popular (w2b; 61) and student writing (w1a; 64) show the lowest PrPf frequencies, while public dialogues (s1b; 119) and letters (w1b; 152) can be found toward the other end of the hierarchy of PrPf-friendliness. NZE also shows the highest value in relation to AusE, its Antipodean sister variety, when the normalized frequencies for the individual text types are correlated, although the correlation is only moderate (r = 0.68).

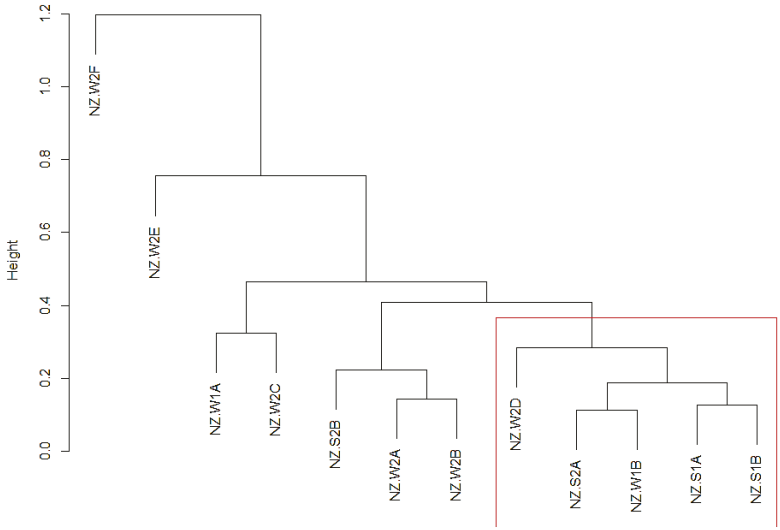


Figure 5.4.17. Cluster dendrogram of similarity across ICE text categories in NZE

In a similar fashion to the similarity analyses of the other varieties, Figure 5.4.17, showing the cluster dendrogram (c = 0.98) for the ICE-NZ data, indicates that three out of the four spoken textual categories align closely toward the right end of the visualization, along with another involved textual category, letters (w1b). Scripted monologues (s2b) cluster with academic (w2a) and popular (w2b) writing. However, the latter cluster is not distant from that containing the other spoken text types and this is further substantiated by a post-hoc test, which identifies the highlighted cluster as significant, while scripted monologues (s2b) are included in that cluster when multiscale bootstrap sampling is applied (not shown).

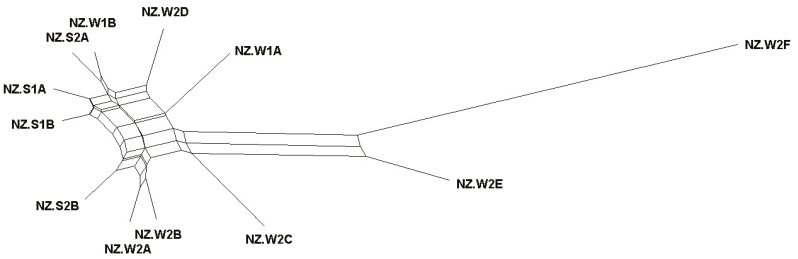


Figure 5.4.18. NeighborNet of similarity across ICE text categories in NZE

Indeed, as indicated in Figure 5.4.18, all four spoken texts cover adjacent nodes in the NeighborNet analysis. It is further illustrated that the majority of text categories groups around a common (imagined) center, which underscores the relative homogeneity of the data, while persuasive (w2e) and, most markedly, creative (w2f) writing occupy distant nodes. Once again, Biber & Conrad's (2009: 261) hypothesis that spoken texts usually align more closely while written texts show more variability seems to bear out for NZE when comparing the distributions of factors determining PrPf usage.

5.5 Canadian English (phase 5)

Canadian English is typically categorized as an L1 variety. Only a brief overview of its multifaceted settlement history can be provided here (see Schneider 2007: 238–250 for a detailed description).[252] Above all, Canada represents a special case due to a strong French presence, which preceded English influence and has left its footprint linguistically (bilingualism/French-speaking regions) and, more strongly, in relation to identity issues. Apart from small-scale earlier attempts, British settlement started

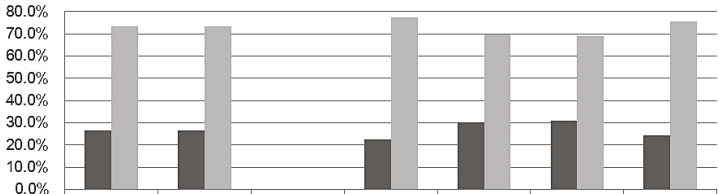
[252] Schneider (2007: 238) also draws attention to the fact that developments in Newfoundland, the island which was settled by people of British origin as early as 1497 but remained a colony of its own until the middle of the 20th century, are to be seen as separate. Political separation and resulting linguistic consequences do no longer apply for PDE, of course.

at the beginning of the 18th century and was boosted by immigration of the American "Loyalists", a group that opposed the independence movement that eventually resulted in the establishment of the US, and further in the 19th century by immigration from the British Isles. While strong ties with Britain have remained, Canada cut the political cord to Britain step by step. In 1867, it gained dominion status and was granted full legal equality in the statute of Westminster in 1931. 1947 then saw the establishment of a separate Canadian citizenship and independence was formally finalized with the 1982 Canada Act (Schneider 2007: 239–247).

Despite these historically strong ties to Britain and the alleged preference for a BrE model (Halford 2002: 32), most observers agree that extralinguistic circumstances are decisive, making AmE (through the "Loyalists") the main input variety to CanE. This development has been reinforced naturally by the geographic proximity to the US, resulting cross-border contacts and the present global influence of AmE, so that CanE now covers a middle position between BrE and AmE (Schneider 1983b: 54–55; Allerton et al. 2002: XII; Hickey 2004: 14) or is even viewed as constrained in its development as a distinctive variety by AmE (Leitner 1992: 202). In contrast, others have argued that the very distinctiveness of CanE is grounded in its unique combination of features from BrE and AmE on all levels of linguistic analysis (Schneider 1983b: 55; Halford 2002: 42), while recent decades have seen the emergence of characteristic phonological patterns such as "Canadian Raising" (Schneider 2007: 244).

Evidence on specific features of CanE grammar is very rare, however (Schneider 2007: 244), and for the area of the PrPf the only statement that could be found in the literature is that, due to Irish influence, *after*-perfects are salient in some areas of Newfoundland (Halford 2002: 42; Kortmann & Szmrecsanyi 2004; see also Section 6.4). Therefore, the working hypothesis is that CanE is similar to AmE in the area of the PrPf, which in turn suggests that it represents a less PrPf-friendly variety and that the distribution of the internal variables may show a distinctive pattern.

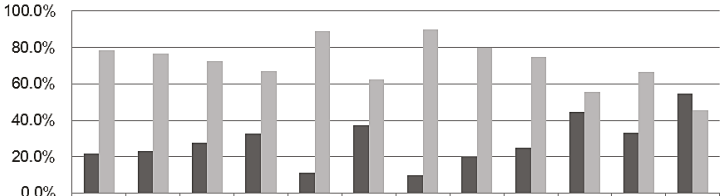
5.5.1 Temporal adverbials



	SPOKEN (n=306)	WRITTEN (n=222)		dialogues (n=151)	monologues (n=155)	non-printed writing (n=74)	printed writing (n=148)
with TA	26.5%	26.6%		22.5%	30.3%	31.1%	24.3%
without TA	73.5%	73.4%		77.5%	69.7%	68.9%	75.7%

Figure 5.5.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-CAN (register and macro-genre differences)

Figure 5.5.1 displays the proportions of PrPf instances in CanE that are specified by a temporal adverbial. The figure indicates that differences between the two registers are marginal and that more than one quarter of all occurrences is specified in both modes of discourse. Although the relative frequencies of specified tokens vary between the macro-genres, none of the differences reaches the 5% level of statistical significance, which suggests a relative homogeneity across the data.



	CAN S1A (n=65)	CAN S1B (n=86)	CAN S2A (n=76)	CAN S2B (n=79)	CAN W1A (n=18)	CAN W1B (n=56)	CAN W2A (n=40)	CAN W2B (n=50)	CAN W2C (n=20)	CAN W2D (n=18)	CAN W2E (n=9)	CAN W2F (n=11)
with TA	21.5%	23.3%	27.6%	32.9%	11.1%	37.5%	10.0%	20.0%	25.0%	44.4%	33.3%	54.5%
without TA	78.5%	76.7%	72.4%	67.1%	88.9%	62.5%	90.0%	80.0%	75.0%	55.6%	66.7%	45.5%

Figure 5.5.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-CAN (detailed text categories)

A different picture emerges when analyzing the detailed ICE text categories. Figure 5.5.2 reveals manifest variance in the distributions across

the textual categories, with specification values ranging from one in ten (academic writing; w2a) to more than half (creative writing; w2f) of all tokens. Note, however, that particularly the result for the latter category should be interpreted with care due to a low token count (11) for this text type, while in general the distribution of the values is much more inconsistent across the written texts.

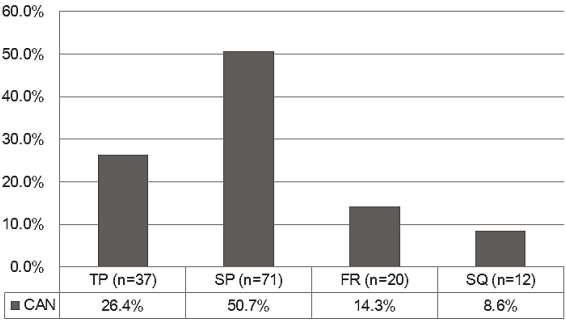
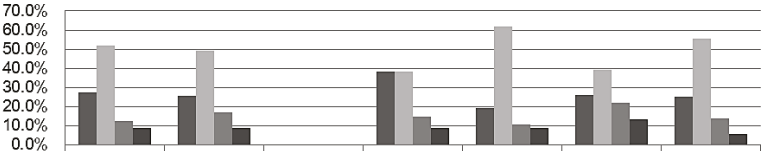


Figure 5.5.3. Types of PrPf adverbial specification: ICE-CAN

Figure 5.5.3 shows that adverbials of span and duration (SP), as in (1), are most frequent and exceed more than half the occurrences that are temporally specified, while adverbials of time-position (TP), as in (2), constitute the second most frequent category with more than one quarter of all instances. The share of frequency adverbials (FR), as in (3), amounts to approximately every seventh occurrence and adverbials of sequence (SQ), as illustrated in (4), are least frequent in the CanE data.

(1) I've been in this limbo *for too long* (ICE-CAN s1a-037)
(2) And I have done absolutely no studying *tonight* (ICE-CAN s1a-097)
(3) when we've let down our hooks *again and again* catching fish every time [...] (ICE-CAN w2f-008)
(4) Half an hour later, I *still* have not seen them. (ICE-CAN w2b-026)



	SPOKEN (n=81)	WRITTEN (n=59)		dialogues (n=34)	monologues (n=47)	non-printed writing (n=23)	printed writing (n=36)
TP	27.2%	25.4%		38.2%	19.1%	26.1%	25.0%
SP	51.9%	49.2%		38.2%	61.7%	39.1%	55.6%
FR	12.3%	16.9%		14.7%	10.6%	21.7%	13.9%
SQ	8.6%	8.5%		8.8%	8.5%	13.0%	5.6%

Figure 5.5.4. Types of PrPf adverbial specification: ICE-CAN (register and macro-genre differences)

Figure 5.5.4 reveals that the relative proportions of the four adverbial categories in the two registers are fairly equal and this impression is confirmed by a Pearson Chi-square test based on the absolute numbers, which yields a non-significant result. In addition, none of the macro-genres deviates significantly from the overall distribution or the distribution of the relevant register, although the relative values seem to imply some variance.

All in all, the results based on ICE-CAN suggest relative homogeneity in this dataset as far as the distribution of the categories of temporal specification by adverbials is concerned. The hierarchy of the different types is SP > TP > FR > SQ, which suggests that PrPfs in CanE, when they are accompanied by a temporal adverbial, typically (at least in 70% of all cases) co-occur with adverbials of the SP and TP type.

5.5.2 Aktionsart

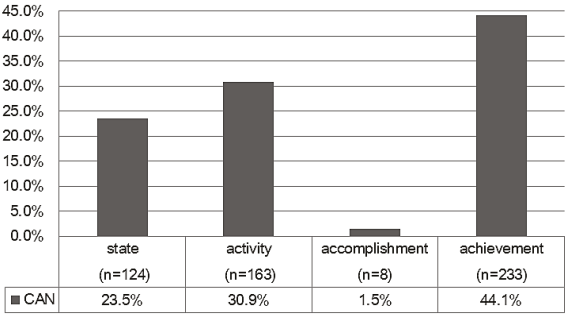


Figure 5.5.5. Aktionsart categories of main verbs in PrPf constructions: ICE-CAN

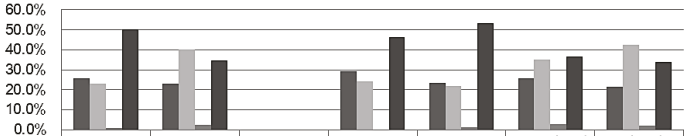
The overall proportions of Aktionsart types in the CanE dataset are presented in Figure 5.5.5. The relative frequency of achievement verbs, as exemplified in (5), exceeds two fifths of all tokens and nearly one third belongs to activities, as in (6). Approximately one quarter is covered by state verbs, as in (7) and accomplishments, as could be expected from the analyses of other varieties, is again a marginal type (see example (8)).

(5) Through the Inter-University Agreement, Quebec universities have *agreed* to permit the transfer of academic credits between them up to a maximum of 6 credits in any one year. (ICE-CAN w2d-003)

(6) Uh already there there's a team of two people in our cegep going around or thinking about this rather and they've *come* to see me once and sort of well you know this is what we're thinking about (ICE-CAN s2a-029)

(7) And what I've suggested is that you look at the books the first the most up-to-date scholarship and uh because the up-to-date stuff some of it has most of it has *included* the earlier stuff so there's no point reinventing the wheel (ICE-CAN s1b-008)

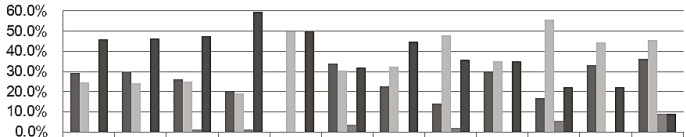
(8) Already this year they've *flown to Winnipeg, Regina and Calgary* [...] (ICE-CAN w1b-013)



	SPOKEN (n=306)	WRITTEN (n=222)		dialogues (n=151)	monologues (n=155)	non-printed writing (n=74)	printed writing (n=148)
state	25.9%	23.0%		29.4%	23.2%	25.7%	21.6%
activity	23.0%	40.1%		24.4%	21.9%	35.1%	42.6%
accomplishment	0.7%	2.3%		0.0%	1.3%	2.7%	2.0%
achievement	50.4%	34.7%		46.2%	53.5%	36.5%	33.8%

Figure 5.5.6. Aktionsart categories of main verbs in PrPf constructions: ICE-CAN (register and macro-genre differences)

It is evident from Figure 5.5.6, which displays the proportions of Aktionsart categories, that speech and writing differ markedly in respect of the relative frequencies of the individual values. This impression is confirmed by a test for statistical significance that contrasts the absolute counts for the two modes of discourse ($\chi^2 = 8.0868$, df = 3, $p < 0.05$, $\varphi_C = 0.112$). The most noticeable deviation from the overall distribution is observable for printed writing, which is the only macro-genre that is significantly different ($\chi^2 = 21.099$, df = 3, $p < 0.001$, $\varphi_C = 0.206$). In contrast, when the four macro-genres are compared with the values for the respective register and when compared to the corresponding genre within the same register, differences in the distributions are not significant, which suggests register-internal homogeneity.



	CAN S1A (n=65)	CAN S1B (n=86)	CAN S2A (n=76)	CAN S2B (n=79)	CAN W1A (n=18)	CAN W1B (n=56)	CAN W2A (n=40)	CAN W2B (n=50)	CAN W2C (n=20)	CAN W2D (n=18)	CAN W2E (n=9)	CAN W2F (n=11)
state	29.2%	29.6%	26.3%	20.3%	0.0%	33.9%	22.5%	14.0%	30.0%	16.7%	33.3%	36.4%
activity	24.6%	24.1%	25.0%	19.0%	50.0%	30.4%	32.5%	48.0%	35.0%	55.6%	44.4%	45.5%
accomplishment	0.0%	0.0%	1.3%	1.3%	0.0%	3.6%	0.0%	2.0%	0.0%	5.6%	0.0%	9.1%
achievement	46.2%	46.3%	47.4%	59.5%	50.0%	32.1%	45.0%	36.0%	35.0%	22.2%	22.2%	9.1%

Figure 5.5.7. Aktionsart categories of main verbs in PrPf constructions: ICE-CAN (detailed text categories)

Considerable variance in the distributions is observable in the fine-grained display according to ICE text types, as shown in Figure 5.5.7. A clear dividing line can be established between the spoken and the written types. The breakdown of the relative frequencies in the former is very homogeneous ($\sigma = 0.02$) apart from the slightly elevated value for achievements in scripted monologues (s2b), while the values are more markedly dispersed in the latter group ($\sigma = 0.07$), where, in addition, achievements do not represent the most frequent category in any text type and a number of cells are empty due to low token counts. Therefore, in order to achieve more conclusive results for the written text types, a larger amount of data would have been desirable, while the findings for the more coarse-grained analysis above can still be considered robust.

5.5.3 Sentence type

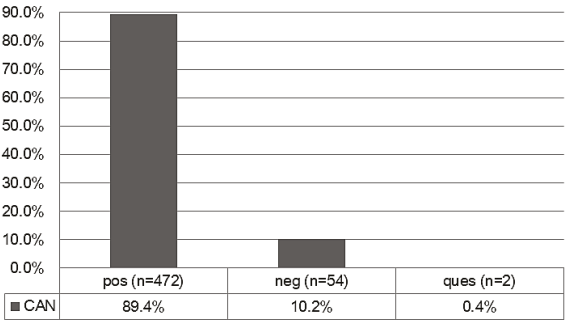
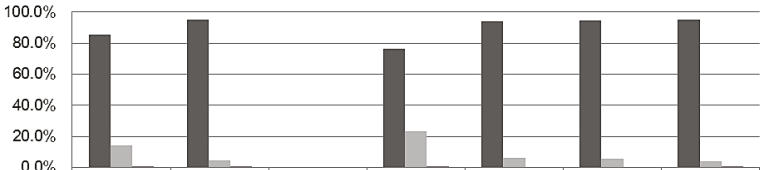


Figure 5.5.8. Distribution of sentence types: ICE-CAN

Figure 5.5.8 shows the distribution of sentence types in the ICE-CAN data. Approximately nine out of ten PrPf tokens can be found in positive statements and the PrPf in negatives, as in (9), or in questions, as in (10), is rare (2 tokens overall).

(9) I'm assuming everything is fine because he hasn't called and I haven't called [...] (ICE-CAN s1b-007)

(10) Are humans innately aggressive, as Konrad Lorenz and other prestigious students of animal behavior have alleged? (ICE-CAN w2a-015)[253]

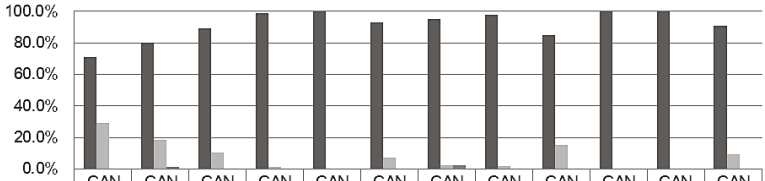


	SPOKEN (n=306)	WRITTEN (n=222)		dialogues (n=151)	monologues (n=155)	non-printed writing (n=74)	printed writing (n=148)
pos	85.3%	95.0%		76.2%	94.2%	94.6%	95.3%
neg	14.4%	4.5%		23.2%	5.8%	5.4%	4.1%
ques	0.3%	0.5%		0.7%	0.0%	0.0%	0.7%

Figure 5.5.9. Distribution of sentence types: ICE-CAN (register and macro-genre differences)

As expected, positives are the most common sentence type across all registers and genres (see Figure 5.5.9). PrPfs in negatives are again three times as frequent in the spoken sections of the ICE-CAN sample, which results in a significant difference between the two modes of discourse ($\chi^2 = 13.687$, df = 2, $p < 0.005$, $\varphi_c = 0.161$). Comparable to a number of other varieties (e.g. AusE and NZE), this difference is due to the high proportion of negatives in dialogues, which approximates one quarter of all PrPf occurrences. The high proportion of negatives also results in the dialogues category being significantly different from its monologues counterpart within the spoken macro-genres ($\chi^2 = 19.997$, df = 2, $p < 0.001$, $\varphi_c = 0.256$), while non-printed and printed writing are similar to one another.

[253] It has to be conceded that this example is not illustrative of a PrPf in a question in the narrow sense as the PrPf does not appear as the main verb.



	CAN S1A (n=65)	CAN S1B (n=86)	CAN S2A (n=76)	CAN S2B (n=79)	CAN W1A (n=18)	CAN W1B (n=56)	CAN W2A (n=40)	CAN W2B (n=50)	CAN W2C (n=20)	CAN W2D (n=18)	CAN W2E (n=9)	CAN W2F (n=11)
pos	70.8%	80.2%	89.5%	98.7%	100.0%	92.9%	95.0%	98.0%	85.0%	100.0%	100.0%	90.9%
neg	29.2%	18.6%	10.5%	1.3%	0.0%	7.1%	2.5%	2.0%	15.0%	0.0%	0.0%	9.1%
ques	0.0%	1.2%	0.0%	0.0%	0.0%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 5.5.10. Distribution of sentence types: ICE-CAN (detailed text categories)

It emerges from Figure 5.5.10 that negatives are absent from a number of written textual categories (student (w1a), instructional (w2d) and persuasive (w2e) writing), while their relative frequency is comparatively high in the dialogic texts (s1a, s1b), unscripted monologues (s2a) and reportage (w2c). As regards the last (written) text type, negatives almost exclusively occur where speech is reported, as is illustrated in (11) to (13).

(11) "The committee is looking at accountability and the committee has been told that previous practices in the office have not been accountable," she said after a committee meeting. (ICE-CAN w2c-010)

(12) The cleanup plans are bogged down in bureaucracy and haven't made real advances in the four years since they began, the IJC says. (ICE-CAN w2c-014)

(13) [...] observes the Reid Group's Bruce Cameron. [...] "Yet the government hasn't taken much heat and that could be because he's one of the few in cabinet who is so respected, nobody has taken much notice of him." (ICE-CAN w2c-017)

While these examples once again nicely exemplify the versatility of written texts in the sense of Biber & Conrad (2009), above all, we can conclude that the distribution of sentence types where PrPfs occur is constrained by register. However, on a deeper level of analysis various pat-

terns also appear when the data are evaluated according to macro-genres and individual ICE text categories.

5.5.4 Semantics

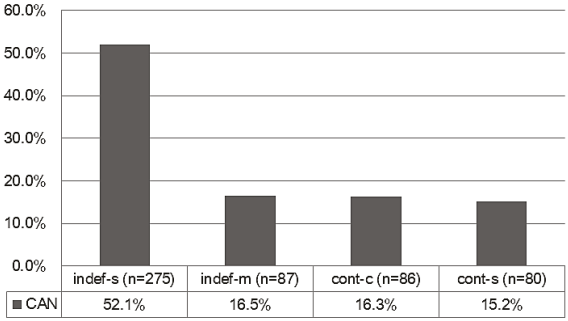


Figure 5.5.11. Distribution of semantic categories: ICE-CAN

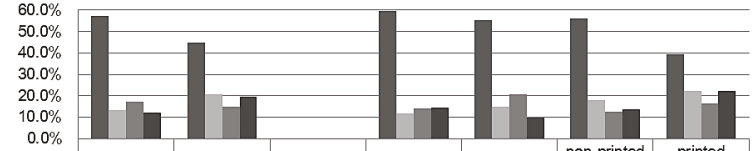
The overall distribution of the semantic categories that can be ascribed to the PrPf in the AusE data is presented in Figure 5.5.11. More than half of all instances can be assigned an indefinite single act/event reading (indef-s), as in (14), while the proportions of the remaining categories, indefinite multiple act/event readings (indef-m), as in (15), continuative acts/events (cont-c), as in example (16), and continuative states (cont-s), as in (17), are nearly equal with approximately one sixth of all PrPf tokens from the sample for each type.

(14) But Dore explained that world women champion Jill Trenary, and the outstanding Soviet pairs and dance teams have decided to duck early-season international assignments to concentrate on domestic, European and world championships. (ICE-CAN w2d-018)

(15) Paul and I have gone for long walks in the forest, looking at the animal tracks and trying to identify all the trees and plants. (ICE-CAN w1b-003)

(16) Since then eighteen other Quebec communities have passed pesticide bylaws although none are as sweeping as the law in Hudson (ICE-CAN s2b-039)

(17) The fact that average life expectancy has increased but maximum life span has not is reflected in the increasing "rectangularization" of survival curves [...]. (ICE-CAN w2a-021)



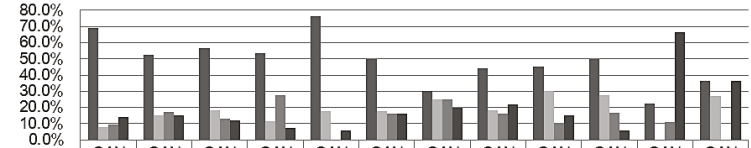
	SPOKEN (n=306)	WRITTEN (n=222)		dialogues (n=151)	monologues (n=155)	non-printed writing (n=74)	printed writing (n=148)
indef-s	57.3%	44.8%		59.6%	55.1%	56.2%	39.2%
indef-m	13.4%	20.8%		11.9%	14.7%	17.8%	22.3%
cont-c	17.3%	14.9%		13.9%	20.5%	12.3%	16.2%
cont-s	12.1%	19.5%		14.6%	9.6%	13.7%	22.3%

Figure 5.5.12. Distribution of semantic categories: ICE-CAN (register and macro-genre differences)

It is apparent from Figure 5.5.12 that differences in the distributions of the categories exist between speech and writing, with notably less indef-s in the latter, and these differences are statistically significant ($\chi^2 = 13.294$, df = 3, $p < 0.005$, $\varphi_C = 0.159$). This significant difference is mainly due to printed writing, which is the only macro-genre that is significantly different from the overall distribution ($\chi^2 = 9.6389$, df = 3, $p < 0.05$, $\varphi_C = 0.119$) but not from non-printed writing or when the values for writing as a whole are taken as a baseline. Note also the small effect size for the comparison with the overall distribution. Otherwise, the semantic readings are distributed fairly similarly across all registers and macro-genres.

Figure 5.5.13 shows that all text types, apart from persuasive writing (w2e), have indef-s as most frequent semantic reading. Some idiosyncratic patterns are observable, however, particularly within the written categories. As is indicated by the SD value, their overall average dispersion ($\sigma = 0.13$) is more than twice the value for the spoken texts ($\sigma = 0.05$). Cont-s is highly frequent in creative (w2f) and persuasive writing (w2e), although the findings for the latter text type, also with regard to the complete absence of indef-m, have to be interpreted cautiously due to the low number of tokens in this category (9). On the con-

trary, comparatively high values are observable for indef-m in reportage (w2c), instructional (w2d) and creative (w2f) writing.

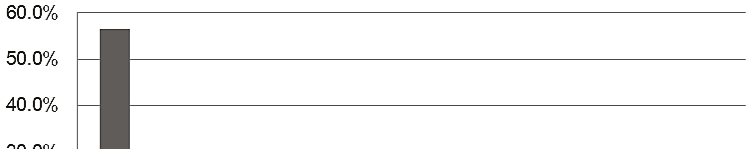


	CAN S1A (n=65)	CAN S1B (n=86)	CAN S2A (n=76)	CAN S2B (n=79)	CAN W1A (n=18)	CAN W1B (n=56)	CAN W2A (n=40)	CAN W2B (n=50)	CAN W2C (n=20)	CAN W2D (n=18)	CAN W2E (n=9)	CAN W2F (n=11)
indef-s	69.2%	52.3%	56.6%	53.8%	76.5%	50.0%	30.0%	44.0%	45.0%	50.0%	22.2%	36.4%
indef-m	7.7%	15.1%	18.4%	11.3%	17.6%	17.9%	25.0%	18.0%	30.0%	27.8%	0.0%	27.3%
cont-c	9.2%	17.4%	13.2%	27.5%	0.0%	16.1%	25.0%	16.0%	10.0%	16.7%	11.1%	0.0%
cont-s	13.8%	15.1%	11.8%	7.5%	5.9%	16.1%	20.0%	22.0%	15.0%	5.6%	66.7%	36.4%

Figure 5.5.13. Distribution of semantic categories: ICE-CAN (detailed text categories)

In conclusion, the ICE-CAN data suggest that the distribution of the individual semantic readings that can be assigned to PrPf occurrences is mainly determined by mode of discourse, while a variety of patterns is in evidence on a more detailed level of analysis, particularly among the informational written ICE text types.

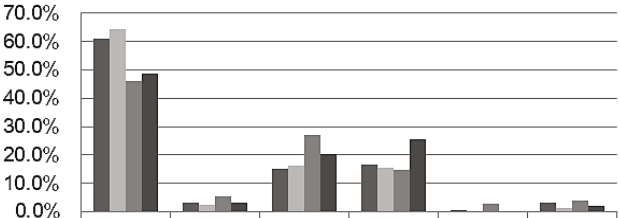
5.5.5 Preceding time reference



	PrTs (n=298)	PrTp (n=18)	SPst (n=98)	Pstp (n=0)	SPrPf (n=98)	PrPfp (n=3)	PaPfs (n=0)	PaPfp (n=0)	F (n=13)
CAN	56.4%	3.4%	18.6%	0.0%	18.6%	0.6%	0.0%	0.0%	2.5%

Figure 5.5.14. TR forms preceding the PrPf: ICE-CAN

Figure 5.5.14 displays the proportions of preceding TR forms in the CanE data. PrPfs that are preceded by PrTs account for more than half of all tokens, and approximately one fifth of occurrences is preceded by SPrPf and SPst respectively. Once again, only PrTp and F are further categories that cover a noticeable share of PrPf instances from the sample.

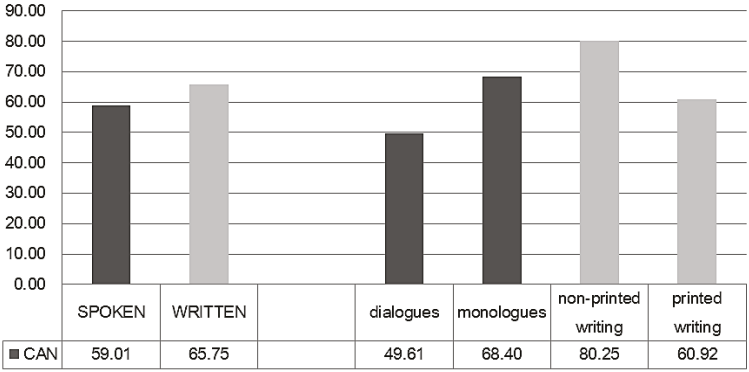


	PrTs (n=298)	PrTp (n=18)	SPst (n=98)	SPrPf (n=98)	PrPfp (n=3)	F (n=13)
dialogues	60.9%	3.3%	15.2%	16.6%	0.7%	3.3%
monologues	64.5%	2.6%	16.1%	15.5%	0.0%	1.3%
non-printed writing	45.9%	5.4%	27.0%	14.9%	2.7%	4.1%
printed writing	48.6%	3.4%	20.3%	25.7%	0.0%	2.0%

Figure 5.5.15. TR forms preceding the PrPf: ICE-CAN (macro-genre differences)

Figure 5.5.15 shows the relative frequencies of preceding TR forms in the individual macro-genres. Forms other than PrTs, SPst and SPrPf occur only sporadically, as could be expected from the analysis of other varieties. In the ICE-CAN sample, speech and writing yield significant differences ($\chi^2 = 12.215$, df = 5, $p < 0.05$, $\varphi_c = 0.152$; relative proportions not shown in Figure 5.5.15), as PrTs is less common in the written sections. Apart from that, the distributions for the macro-genres are comparable (i.e. not significantly different) when contrasted genre-internally, with the breakdown for the registers and also with the overall absolute counts. This indicates that the distribution of preceding TR forms in the CanE data is relatively homogeneous across the board, while mode of discourse exerts some influence.

5.5.6 Perfect-friendliness and text type effects



	SPOKEN	WRITTEN		dialogues	monologues	non-printed writing	printed writing
CAN	59.01	65.75		49.61	68.40	80.25	60.92

Figure 5.5.16. Frequency of PrPf occurrences in ICE-CAN registers and macro-genres (normalized frequency per 10,000 words)

It was suggested above that CanE can be found on the less PrPf-friendly side of the hierarchy of varieties and indeed this is the case. The overall average for ICE-CAN (64 PrPfs/pttw) is below the average across all varieties (69), and the average normalized frequency of the corpus components is lower in the spoken than in the written sections (see Figure 5.5.16). Non-printed writing is the macro-genre with the highest PrPf frequency (80).

As far as the dispersion of the normalized PrPf frequencies is concerned, a value of $\sigma = 20.02$ can be established for the dataset. PrPfs are least common in creative writing (w2f; 25) and private dialogues (s1a; 42), while the highest normalized frequencies occur in scripted monologues (s2b; 84) and letters (w1b; 106). With regard to the correlation of the values for the ICE text categories, PhiE appears most strongly associated with CanE ($r = 0.75$), which indicates that the relationship between the two through the parent variety AmE may play a part here.

The hierarchical cluster dendrogram ($c = 0.92$) of the CanE data in Figure 5.5.17 shows that the spoken text categories either cluster together directly (public dialogues (s1b) and unscripted monologues (s2a)) or can be found in neighboring clusters (private dialogues (s1a) and scripted monologues (s2b)). A post-hoc test shows the highlighted cluster

containing the latter pair to be significant. Again, letters (w1b) as another involved text type appears nearby.

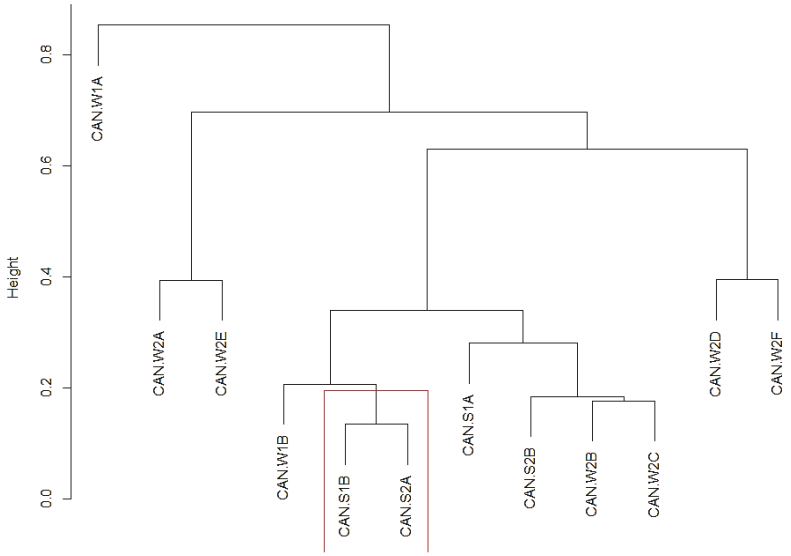


Figure 5.5.17. Cluster dendrogram of similarity across ICE text categories in CanE

The picture becomes clearer in the non-hierarchical display in Figure 5.5.18. None of the spoken text categories are very distant from each other and group toward the bottom left of the NeighborNet with a number of the written text types. Five (all written) out of the twelve categories, however, appear relatively apart. Most notably, student writing (w1a), which also contrasts with all the remaining text types in the dendrogram above, although the NeighborNet suggests much less distance between student and instructional (w2d) writing, for instance, compared to persuasive writing (w2e) and thus offers a more insightful analysis. In sum, the CanE data relying on the distributions of variables that determine the PrPf also support Biber & Conrad's (2009: 261) idea that spoken texts usually align more closely with regard to their structural features while written texts are more variable.

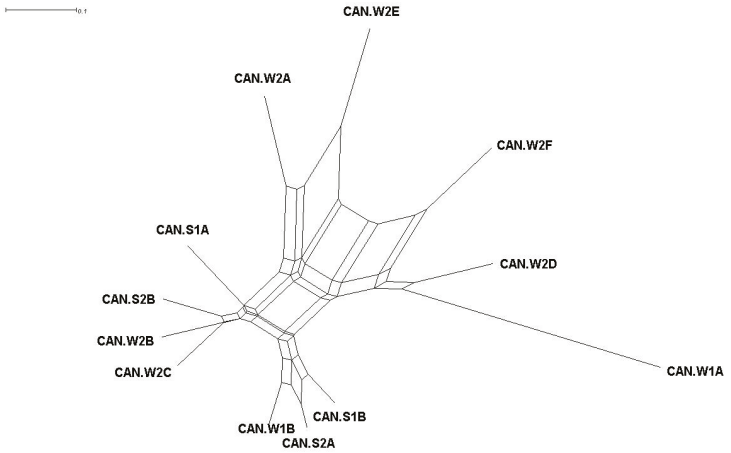


Figure 5.5.18. NeighborNet of similarity across ICE text categories in CanE

5.6 Philippine English (phase 3–4)

In contrast to the other varieties discussed so far, the primary ancestor of PhiE is AmE. At the same time, the Philippines can be seen as a typical exploitation colony that underwent colonial rule by various countries. After more than 300 years of Spanish rule, the Philippines were ceded to the USA at the turn to the 20th century. The US declared English the official language of the country and enacted strong language planning measures such as introducing American teachers[254] and making English the language of instruction (Schneider 2007: 140). The Philippines subsequently became semi-independent in 1935, when the country was granted Commonwealth status and its own government and, after a phase of Japanese occupation from 1942 to 1945, fully independent in 1946 with the establishment of the Republic of the Philippines.

English has continued its powerful presence, although successive governments since 1937 have advocated a bilingual language policy with Tagalog (later rebranded "Filipino"; Görlach 2002: 111) as national lan-

[254] The majority of English teachers, however, has always been represented by native Filipinos (Kachru & Nelson 2006: 184–185, 187), so indigenizing forces most likely have been operating for an extended period.

guage. However, the implementation of Filipino has not ousted English. The present situation can be seen as diglossic with English as the language of business as well as public and formal contexts, while Filipino, local varieties or "Taglish" cover more personal and day-to-day domains. According to Sibayan (1996), proficiency in English is very strongly determined by socioeconomic class and the often corresponding educational level. Schneider (2007: 142) observes that this situation results in "ambivalent attitudes" toward English as a medium of communication that is on the one hand respected, particularly for its economic value, but on the other hand still scorned for its association with colonialism and political elitism.[255] Therefore, it has been hypothesized that whether English presence will continue depends very much on the future development of Taglish and especially Filipino also for professional and intellectual domains (Thompson 2003: 265).[256]

For historical reasons, PhiE is commonly viewed as close to AmE in pronunciation (Wächtler 1977: 64) and also in other linguistic areas (Köppl 1983b: 126). Characteristic indigenized features are mainly lexical but also lexicogrammatical (Schneider 2007: 142) and a number of observations on tense usage in PhiE can be found in the literature. Generally speaking, due to substrate influence, speakers of PhiE focus on aspectual distinctions "in terms of the beginning and completion of actions" (Kachru & Smith 2008: 93; see also Kachru & Nelson 2006: 191) rather than tense, and often, this distinction is marked analytically (e.g. with temporal adverbials such as *already*) without any or only reduced tense/aspect marking (Kachru & Nelson 2006: 42).[257]

---

[255] Cf. Sibayan (1996: 144–145). See further Thompson (2003) for an in-depth sociolinguistic analysis and Manarpaac (2003) for a study of the Philippine language policy.

[256] Schneider notes that the situation for PhiE is special as, despite showing some signs of phase 4 of his dynamic model (such as national literature in English and attempts at codification; see also Bautista 2008: 217), the developmental circle has halted or "fossilized" (2007: 58) due to external circumstances (2007: 143) and English is about to lose (Manarpaac 2003: 479–481) or, as others claim (Sibayan 1996: 144), has *de facto* already lost its co-official status.

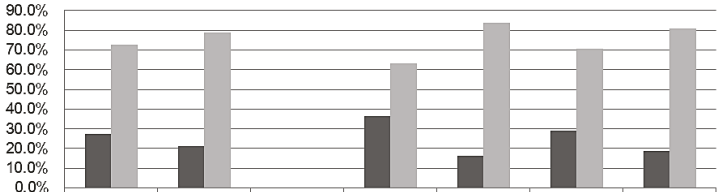
[257] Using adverbials without any further grammatical marking to convey tense is also common in pidgin languages (Labov 1990: 16). Alternatively, it is conceivable that it represents a general learner feature (Bardovi-Harlig 1997: 377; see also Section 6.4).

With regard to the PrPf in particular, Sibayan (1996: 161) notes that many speakers of PhiE do not reach an AmE/BrE target-like usage, while an apparent-time study, which has also shown that such a PrPf usage is mastered only by few speakers, has revealed that a "decline" in "correct" PrPf usage has been observable diachronically and is especially prominent among younger generations (Gonzales et al. 2003: 54, 110, 115). This finding could be reinterpreted such that a kind of nativized usage is emerging, with Past Perfect use instead of the PrPf (see example (1)) as another characteristic feature (McArthur 2002: 346; Trudgill & Hannah 2008: 143).

(1) the President *had accumulated* unexplained wealth and this is indicated with the amounts in the accounts (ICE-PHI s1b-061)

Therefore, the working hypotheses for PhiE are, first, that due to its orientation toward AmE it will be less PrPf-friendly and, second, that PrPf forms that are unmarked or only partly marked may occur and that speakers will use more temporal adverbials as a mechanism of compensation with these forms (particularly in spoken and involved texts) and potentially also with the PrPf in general.

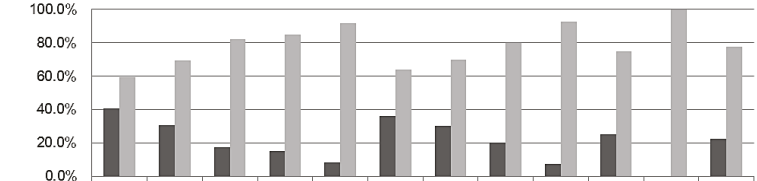
5.6.1 Temporal adverbials



	SPOKEN (n=320)	WRITTEN (n=207)		dialogues (n=172)	monologues (n=148)	non-printed writing (n=48)	printed writing (n=159)
with TA	27.2%	21.3%		36.6%	16.2%	29.2%	18.9%
without TA	72.8%	78.7%		63.4%	83.8%	70.8%	81.1%

Figure 5.6.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-PHI (register and macro-genre differences)

Figure 5.6.1 displays the proportions of PrPf instances in the ICE-PHI data that are specified by a temporal adverbial. The differences between the values of the spoken and the written part are not significant, although the relative values, with more than one fifth of all tokens for the written and more than one quarter for the spoken part, differ markedly. The spoken macro-genres yield the most extreme values in both directions and both dialogues ($\chi^2 = 8.9597$, df = 1, $p < 0.005$, $\varphi_c = 0.113$) and monologues ($\chi^2 = 4.8777$, df = 1, $p < 0.05$, $\varphi_c = 0.085$) are significantly different from both the overall value and naturally very strongly from each other ($\chi^2 = 16.743$, df = 1, $p < 0.001$, $\varphi_c = 0.229$). The distribution of the values in the written genres, in contrast, is internally relatively homogeneous, although the relative values yield higher proportions in the involved (non-printed) section.



	PHI S1A (n=103)	PHI S1B (n=69)	PHI S2A (n=68)	PHI S2B (n=80)	PHI W1A (n=12)	PHI W1B (n=36)	PHI W2A (n=30)	PHI W2B (n=55)	PHI W2C (n=27)	PHI W2D (n=24)	PHI W2E (n=14)	PHI W2F (n=9)
with TA	40.8%	30.4%	17.6%	15.0%	8.3%	36.1%	30.0%	20.0%	7.4%	25.0%	0.0%	22.2%
without TA	59.2%	69.6%	82.4%	85.0%	91.7%	63.9%	70.0%	80.0%	92.6%	75.0%	100.0%	77.8%

Figure 5.6.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-PHI (detailed text categories)

As could be expected from the previous analysis, considerable variance in the proportions of PrPfs that are temporally specified is evident in the different text types (see Figure 5.6.2). High proportions (> 30%) of tokens with temporal adverbials occur in dialogues (s1a and s1b) and furthermore in letters (w1b) and academic writing (w2a). Apart from the latter text category, these represent the more involved and interactive types and the above hypothesis that in these types more specification occurs is borne out by the ICE-PHI data, suggesting that register and text type effects clearly play a role in this variety.

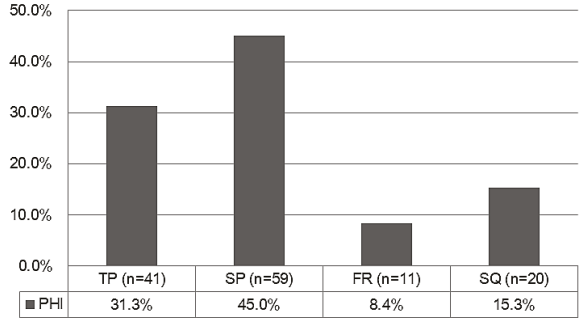


Figure 5.6.3. Types of PrPf adverbial specification: ICE-PHI

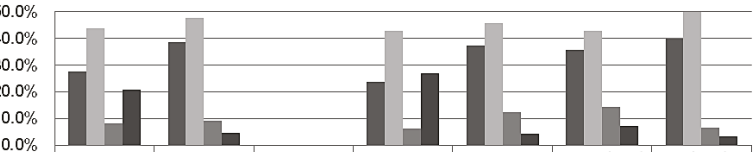
Figure 5.6.3 displays the distribution of the types of temporal specification of PrPfs by an adverbial in PhiE. Adverbials of span and duration (SP), as exemplified in (2), cover more than two fifths of all PrPfs and represent the most frequent category. They are followed by adverbials of time-position (TP), as in (3), with nearly one third of all instances and adverbials of sequence (SQ), as in (4). Frequency adverbials (FR), as in example (5), are rare.

(2) We will not be where we are today but for the faith and spirit that *since June twelve eighteen ninety-eight* has fortified us in trials and crises that sometimes seem overwhelming (ICE-PHI s2b-028)

(3) We are unlocking key bottlenecks that have hampered the use of O D A *in the past year* (ICE-PHI s2b-049)

(4) Well I I I think it's uh primarily because of uh the the structural reforms needed to uh to provide uh a a a transparent playing field for all investors uh in the country has not really seen it its its fruition *yet* (ICE-PHI s1b-049)

(5) Government reorganization has occurred *five times in the post-independence period*, an average of one reorganization every eight years (Cola 1993:382). (ICE-PHI w2a-019)



	SPOKEN (n=87)	WRITTEN (n=44)		dialogues (n=63)	monologues (n=24)	non-printed writing (n=14)	printed writing (n=30)
TP	27.6%	38.6%		23.8%	37.5%	35.7%	40.0%
SP	43.7%	47.7%		42.9%	45.8%	42.9%	50.0%
FR	8.0%	9.1%		6.3%	12.5%	14.3%	6.7%
SQ	20.7%	4.5%		27.0%	4.2%	7.1%	3.3%

Figure 5.6.4. Types of PrPf adverbial specification: ICE-PHI (register and macro-genre differences)

The relative proportions of the types of adverbial specification for the two modes of discourse as displayed in Figure 5.6.4 show some, though non-significant, variance. The same applies to the differences in the distributions for the four macro-genres, where no significant differences emerge from the overall distribution, the distribution of the corresponding register and also genre-internally. The big differences in the relative values can be explained by comparatively small sample sizes for some of the genres (e.g. 14 tokens for non-printed writing).

On the basis of the ICE-PHI data, we can establish an overall hierarchy of types of temporal specification that typically occur with the PrPf in PhiE along the lines SP > TP > SQ > FR. One noteworthy finding is the high proportion of SQ adverbials (mostly *already, yet, still*) in speech and in dialogues in particular, which is higher than in the remaining macro-genres and also higher in dialogues than in all the varieties considered so far by at least one third on average. A possible explanation for this outcome is that the compensatory use of these adverbials as markers of completion (see above and example (6) with *yet* in non-final position) may also be generalized in non-vernacular data such as contained in ICE, and might thus be seen as additional markers of the perfect in PhiE that regularly co-occur with the PrPf. This phenomenon seems most likely to surface in dialogues, as the most involved and interactive category.

(6) He has not he has not finished *yet* his college (ICE-PHI s1a-021)

5.6.2 Aktionsart

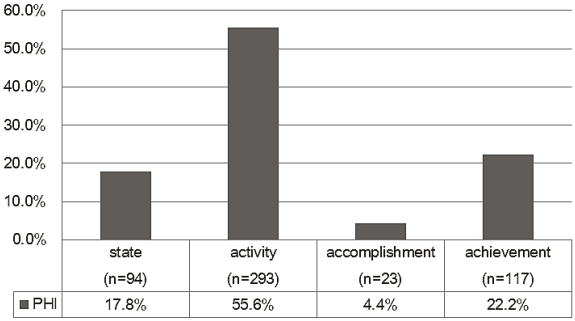
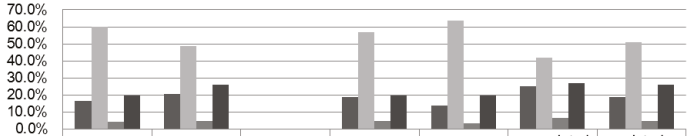


Figure 5.6.5. Aktionsart categories of main verbs in PrPf constructions: ICE-PHI

Figure 5.6.5 displays the overall proportions of Aktionsart categories in the ICE-PHI data. Activity verbs, as in (7), covering more than half of all PrPf tokens, are the most frequent type, followed by achievements, as exemplified in (8). State verbs, as in (9), occur in more than one quarter of all instances and accomplishments, as in (10), are again the least common.

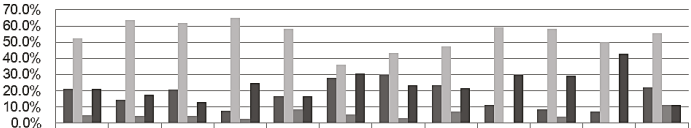
(7) This is what God has *called* us to do hija. (ICE-PHI w2f-006)
(8) THOSE who have *given up* on government should take heart. (ICE-PHI w2e-002)
(9) Men have *remained* even ignorant to the fact that pollutants added to the air are reactants in reactions which are taking place in our atmosphere. (ICE-PHI w2a-035)
(10) And ever since that night I I I've *become* the happiest girl here you know [...] (ICE-PHI s1a-019)



	SPOKEN (n=320)	WRITTEN (n=207)		dialogues (n=172)	monologues (n=148)	non-printed writing (n=48)	printed writing (n=159)
state	16.3%	20.3%		18.6%	13.5%	25.0%	18.9%
activity	60.0%	48.8%		57.0%	63.5%	41.7%	50.9%
accomplishment	4.1%	4.8%		4.7%	3.4%	6.3%	4.4%
achievement	19.7%	26.1%		19.8%	19.6%	27.1%	25.8%

Figure 5.6.6. Aktionsart categories of main verbs in PrPf constructions: ICE-PHI (register and macro-genre differences)

There are mixed results with regard to the homogeneity of the data. The differences in the distributions of Aktionsart categories between the two registers, as shown in Figure 5.6.6, are not significant and printed writing is the only macro-genre in which the 5% level of statistical significance is reached compared with the overall distribution (χ^2= 7.8497, df = 3, $p < 0.05$, $\varphi_C = 0.108$). While none of the remaining genres is significantly different from the overall distribution and the same applies for all macro-genres when compared with the distribution of their corresponding registers, it emerges that, register-internally, the proportions are not homogeneous. Both dialogues and monologues (χ^2 = 9.8962, df = 3, $p < 0.05$, $\varphi_C = 0.173$) and also non-printed and printed writing (χ^2 = 11.714, df = 3, $p < 0.01$, $\varphi_C = 0.236$) differ significantly when compared to each other, meaning that genre differences are salient in PhiE for this variable.



	PHI S1A (n=103)	PHI S1B (n=69)	PHI S2A (n=68)	PHI S2B (n=80)	PHI W1A (n=12)	PHI W1B (n=36)	PHI W2A (n=30)	PHI W2B (n=55)	PHI W2C (n=27)	PHI W2D (n=24)	PHI W2E (n=14)	PHI W2F (n=9)
state	21.4%	14.5%	20.6%	7.5%	16.7%	27.8%	30.0%	23.6%	11.1%	8.3%	7.1%	22.2%
activity	52.4%	63.8%	61.8%	65.0%	58.3%	36.1%	43.3%	47.3%	59.3%	58.3%	50.0%	55.6%
accomplishment	4.9%	4.3%	4.4%	2.5%	8.3%	5.6%	3.3%	7.3%	0.0%	4.2%	0.0%	11.1%
achievement	21.4%	17.4%	13.2%	25.0%	16.7%	30.6%	23.3%	21.8%	29.6%	29.2%	42.9%	11.1%

Figure 5.6.7. Aktionsart categories of main verbs in PrPf constructions: ICE-PHI (detailed text categories)

Figure 5.6.7 shows the detailed view according to ICE text categories, which clearly yields some variance, although activities are the most frequent and accomplishments the least frequent category across the board. While the distributions in the spoken text types are relatively uniform ($\sigma = 0.03$), achievements are more prominent in the written (and there, particularly in informational) texts, representing the second most common category in letters (w1b), reportage (w2c), instructional (w2d) and persuasive (w2e) writing, resulting in a dispersion that is twice the value for the spoken text types. Note, however, that the findings for some of the written text categories have to be taken with a pinch of salt due to low token counts.

5.6.3 Sentence type

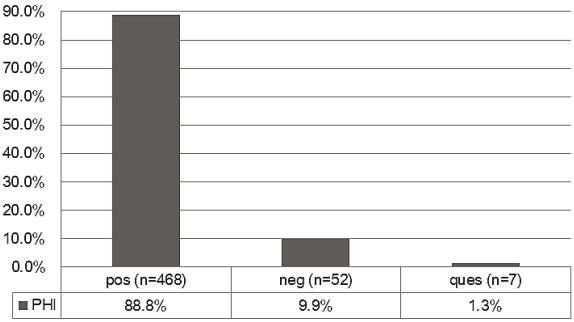
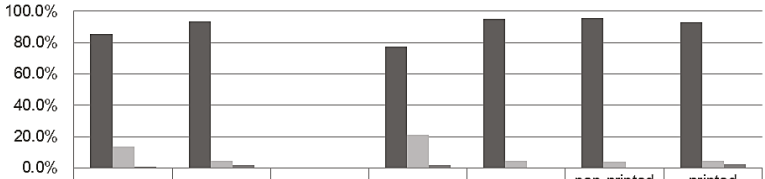


Figure 5.6.8. Distribution of sentence types: ICE-PHI

Also in the PhiE data, nearly nine out of ten PrPf tokens occur in positive statements (see Figure 5.6.8). The PrPf with negatives, as in example (11), accounts for the remaining instances, apart from seven tokens categorized as questions, as in (12).

(11) I am having a lot of trouble focusing on work and studies these days because the issue of the NGEC has not stopped haunting me. (ICE-PHI w1a-002)

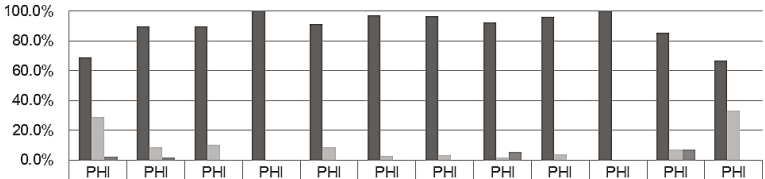
(12) And you are so you've been away for how many years (ICE-PHI s1a-010)



	SPOKEN (n=320)	WRITTEN (n=207)		dialogues (n=172)	monologues (n=148)	non-printed writing (n=48)	printed writing (n=159)
pos	85.6%	93.7%		77.3%	95.3%	95.8%	93.1%
neg	13.4%	4.3%		20.9%	4.7%	4.2%	4.4%
ques	0.9%	1.9%		1.7%	0.0%	0.0%	2.5%

Figure 5.6.9. Distribution of sentence types: ICE-PHI (register and macro-genre differences)

Figure 5.6.9 shows that positives are slightly more frequent in writing, while negatives clearly are more common in speech in ICE-PHI. In consequence, differences in the proportions of the sentence types between the two registers are significant (χ^2 = 12.389, df = 2, p < 0.005, φ_c = 0.152). Once again, the macro-genre of dialogues has the highest rate of negatives, exceeding one fifth of all occurrences, while the proportions for this type are similar in the remaining ones. As a result, dialogues are significantly different both from the overall distribution (χ^2 = 14.751, df = 2, p < 0.001, φ_c = 0.145) and also register-internally from monologues (χ^2 = 21.110, df = 2, p < 0.001, φ_c = 0.257). In contrast, the written macro-genres are fairly homogeneous.



	PHI S1A (n=103)	PHI S1B (n=69)	PHI S2A (n=68)	PHI S2B (n=80)	PHI W1A (n=12)	PHI W1B (n=36)	PHI W2A (n=30)	PHI W2B (n=55)	PHI W2C (n=27)	PHI W2D (n=24)	PHI W2E (n=14)	PHI W2F (n=9)
pos	68.9%	89.9%	89.7%	100.0%	91.7%	97.2%	96.7%	92.7%	96.3%	100.0%	85.7%	66.7%
neg	29.1%	8.7%	10.3%	0.0%	8.3%	2.8%	3.3%	1.8%	3.7%	0.0%	7.1%	33.3%
ques	1.9%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	5.5%	0.0%	0.0%	7.1%	0.0%

Figure 5.6.10. Distribution of sentence types: ICE-PHI (detailed text categories)

Figure 5.6.10 reveals that the PrPf in questions only surfaces in a limited set of textual categories. Not surprisingly, these are dialogues (s1a and s1b) and instructional writing (w2b), where rhetorical questions also occur (see example (13)).

(13) Or aren't we really rather well off considering what we've been through? (ICE-PHI w2b-010)

The comparatively high relative frequency of questions in persuasive writing (w2e) is misleading as the sample for this category is relatively small (14 tokens in total). The only example of a question involving a PrPf is again of a rhetorical nature (see example (14)):

(14) Didn't we say before that we have had enough of old-style politics? (ICE-PHI w2e-006)

While at least 85% of all tokens are positives in the majority of text types, negatives are strongly represented in private dialogues (s1a) and creative writing (w2f). Note, however, that in contrast to some other varieties, the examples from the creative writing section of ICE-PHI do not simulate dialogues.

In sum, some effects of register and macro-genre and, furthermore, individual text type on the proportions of sentence types when PrPfs are present can be observed in ICE-PHI.

5.6.4 Semantics

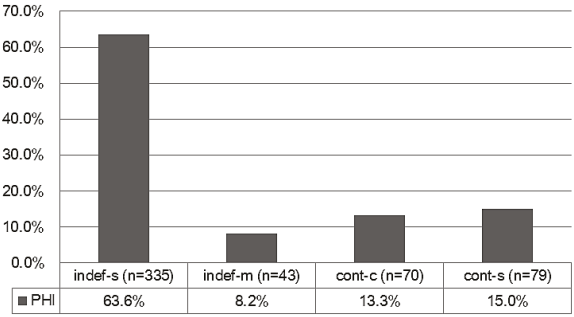


Figure 5.6.11. Distribution of semantic categories: ICE-PHI

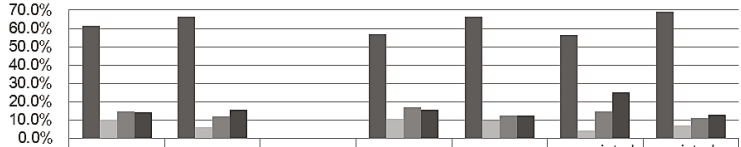
The overall relative frequencies of semantic categories of the PrPf in ICE-PHI are displayed in Figure 5.6.11. Indefinite single acts/events (indef-s), as in example (15), exceed three fifths of all tokens, while the continuative categories acts/events (cont-c), as in (16), and states (cont-s), as in (17), each cover approximately one in seven occurrences. The remaining tokens can be assigned an indefinite multiple act/event reading (indef-m), as in example (18).

(15) At the same time Cojuangc admits that former First Lady Imelda Marcos' entry into the political derby has affected his own presidential bid [...] (ICE-PHI s2b-010)

(16) With no relief in sight, he decided to try Alternative Medicine last August and since then has been on acupuncture sessions twice a week. (ICE-PHI w1b-006)

(17) The House did not touch the debt service allocation although it has been Congress' common practice to cut this and realign the funds to the legislators' pet projects. (ICE-PHI w2c-010)

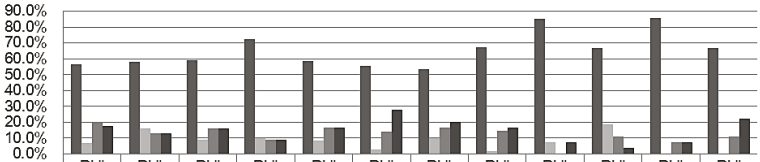
(18) Actually we've watched a lot of movie together right (ICE-PHI s1a-030)



	SPOKEN (n=320)	WRITTEN (n=207)		dialogues (n=172)	monologues (n=148)	non-printed writing (n=48)	printed writing (n=159)
indef-s	61.3%	66.2%		57.0%	66.2%	56.3%	69.1%
indef-m	10.0%	6.2%		10.5%	9.5%	4.2%	6.8%
cont-c	14.7%	11.9%		16.9%	12.2%	14.6%	11.1%
cont-s	14.1%	15.7%		15.7%	12.2%	25.0%	13.0%

Figure 5.6.12. Distribution of semantic categories: ICE-PHI (register and macro-genre differences)

Some minor variance exists between the two modes of discourse with regard to the distribution of semantic categories of PrPfs, as Figure 5.6.12 shows. However, a Pearson Chi-square test reveals that differences are not significant. The proportions of the semantic readings across the macro-genres are also fairly homogeneous and none deviates significantly from the overall distribution or the corresponding register values, nor when compared register-internally.



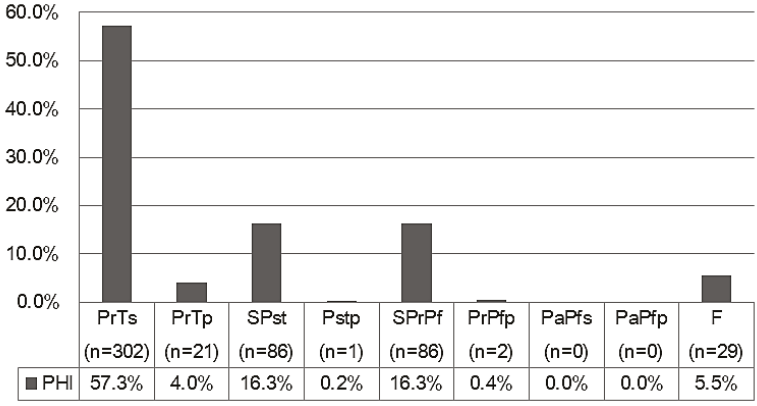
	PHI S1A (n=103)	PHI S1B (n=69)	PHI S2A (n=68)	PHI S2B (n=80)	PHI W1A (n=12)	PHI W1B (n=36)	PHI W2A (n=30)	PHI W2B (n=55)	PHI W2C (n=27)	PHI W2D (n=24)	PHI W2E (n=14)	PHI W2F (n=9)
indef-s	56.3%	58.0%	58.8%	72.5%	58.3%	55.6%	53.3%	67.3%	85.2%	66.7%	85.7%	66.7%
indef-m	6.8%	15.9%	8.8%	10.0%	8.3%	2.8%	10.0%	1.8%	7.4%	18.5%	0.0%	0.0%
cont-c	19.4%	13.0%	16.2%	8.8%	16.7%	13.9%	16.7%	14.5%	0.0%	11.1%	7.1%	11.1%
cont-s	17.5%	13.0%	16.2%	8.8%	16.7%	27.8%	20.0%	16.4%	7.4%	3.7%	7.1%	22.2%

Figure 5.6.13. Distribution of semantic categories: ICE-PHI (detailed text categories)

Figure 5.6.13 shows that indef-s is the majority semantic reading across the board. The written text types show twice as much variance compared to the spoken ones ($\sigma = 0.08$ vs. $\sigma = 0.04$, respectively). The text types letters (w1b) and creative writing (w2f) have high proportions of cont-s. The results for the latter text type should not be overinterpreted due to

the low number of tokens in this category (9), also regarding the lack of tokens for the indef-m category. Further outliers are the high relative frequency of indef-m readings in instructional writing (w2d) and public dialogues (s1b) and of cont-c readings in the text categories private dialogues (s1a) and student (w1a) and academic (w2a) writing. All in all, however, a fairly uniform picture regarding the distribution of semantic categories emerges for PrPf constructions in ICE-PHI, naturally with some exceptions in the detailed text view.

5.6.5 Preceding time reference



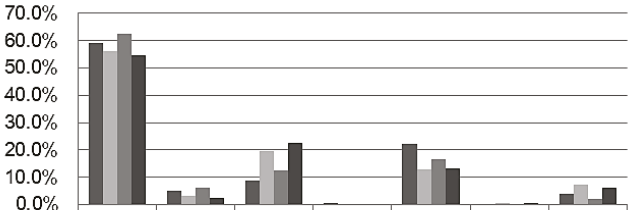
	PrTs (n=302)	PrTp (n=21)	SPst (n=86)	Pstp (n=1)	SPrPf (n=86)	PrPfp (n=2)	PaPfs (n=0)	PaPfp (n=0)	F (n=29)
PHI	57.3%	4.0%	16.3%	0.2%	16.3%	0.4%	0.0%	0.0%	5.5%

Figure 5.6.14. TR forms preceding the PrPf: ICE-PHI

Figure 5.6.14 presents the proportions of preceding TR forms in ICE-PHI. PrTs, with approximately three out of five, is the most common variant, while SPrPf and SPst each precede nearly every sixth PrPf occurrence. Minor noteworthy variants in the data sample are preceding F and PrTp.

The macro-genre analysis presented in Figure 5.6.15 also reveals that PrTs, SPst and SPrPf are the only preceding TR classes with a notable share. The differences in the proportions of the spoken and written sections (relative values not shown in Figure 5.6.15) fail to reach statistical significance at the 5% level and the distributions of the macro-genres are also uniform. The only genre effect that yields a significant result

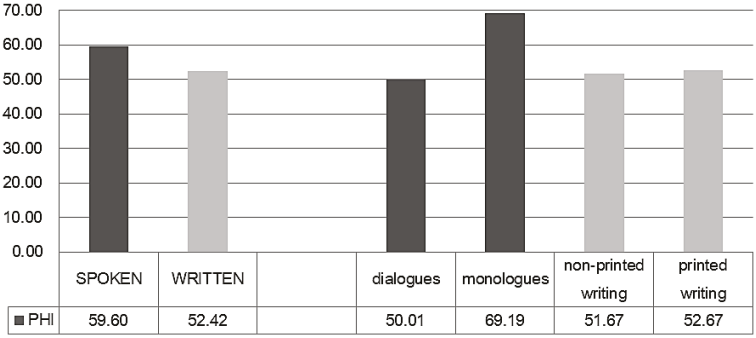
occurs when dialogues are contrasted with monologues (χ^2 = 15.056, df = 6, $p < 0.05$, $\varphi_c = 0.217$), which suggests that, overall, the distribution of TR forms preceding PrPfs in PhiE is fairly homogeneous, apart from some genre-internal differences in speech.



	PrTs (n=302)	PrTp (n=21)	SPst (n=86)	Pstp (n=1)	SPrPf (n=86)	PrPfp (n=2)	F (n=29)
dialogues	59.3%	5.2%	8.7%	0.6%	22.1%	0.0%	4.1%
monologues	56.1%	3.4%	19.6%	0.0%	12.8%	0.7%	7.4%
non-printed writing	62.5%	6.3%	12.5%	0.0%	16.7%	0.0%	2.1%
printed writing	54.7%	2.5%	22.6%	0.0%	13.2%	0.6%	6.3%

Figure 5.6.15. TR forms preceding the PrPf: ICE-PHI (macro-genre differences)

5.6.6 Perfect-friendliness and text type effects



	SPOKEN	WRITTEN		dialogues	monologues	non-printed writing	printed writing
PHI	59.60	52.42		50.01	69.19	51.67	52.67

Figure 5.6.16. Frequency of PrPf occurrences in ICE-PHI registers and macro-genres (normalized frequency per 10,000 words)

As was indicated above, PhiE, due to its close relation to AmE, commonly is conceived as a variety that overall is not PrPf-friendly. Indeed, with

approximately 55 PrPfs/pttw, the ICE-PHI data yield the second lowest PrPf frequency across the board. PrPfs are slightly more common in speech, but the macro-genre view reveals that the values for dialogues (50) and monologues (69) diverge considerably and constitute the extreme poles, while the values for the two written genres are almost identical (see Figure 5.6.16).

The ICE-PHI data are internally homogeneous, with the lowest variance of normalized PrPf frequencies (σ = 16.63). PrPfs are rarest in creative (w2f; 22) and academic (w2a; 39) writing and most frequent in scripted monologues (s2b; 83) and popular writing (w2b; 74). The closest variety in terms of correlation of the normalized PrPf frequencies is IrE (r = 0.82), a finding that seems naturally plausible as the latter variety is also traditionally less PrPf-friendly for the reasons explained above (see Section 5.2).

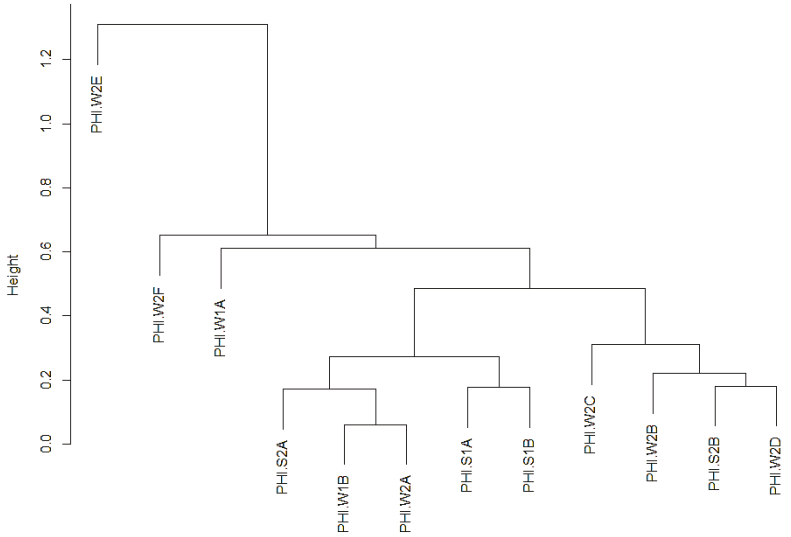


Figure 5.6.17. Cluster dendrogram of similarity across ICE text categories in PhiE

Figure 5.6.17, which displays the hierarchical cluster dendrogram for the ICE-PHI data (c = 0.96), shows that for these data, the clustering of the spoken text categories is not as unambiguous as in other varieties. How-

ever, public (s1b) and private (s1a) dialogues cover adjacent leaves and unscripted monologues (s2a) cluster at the next level, while scripted monologues (s2b) represents an outlier. Letters (w1b) again can be found in proximity to three out of the four spoken text types. Yet, no significant spoken or involved (*sensu* Biber & Conrad 2009) cluster emerges.

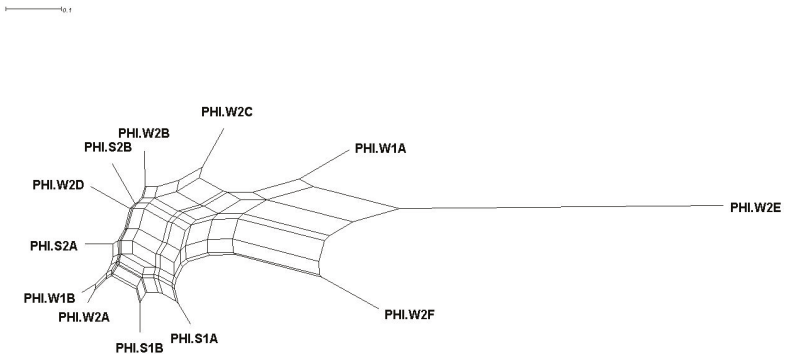


Figure 5.6.18. NeighborNet of similarity across ICE text categories in PhiE

The non-hierarchical NeighborNet view in Figure 5.6.18 reveals further details. First, it emerges that the majority of text types can be found in closely neighboring nodes toward the left of the network representation, while many splits occur. Second, scripted (s2b) and unscripted (s2a) monologues in this view appear closer than could have been inferred from the hierarchical view above, while the former does not share as many splits with the remaining spoken categories as the latter. Third, as could be expected from previous analyses, a number of text types can be categorized as outliers from an imagined core of ICE-PHI texts. These are persuasive (w2e) and, less markedly, student (w1a) and creative (w2f) writing. Again, the ICE-PHI data also provide some indication for the similarity of spoken texts in the area of the PrPf and thus lend further support to the idea that spoken texts are more uniform when compared

to written ones (Biber & Conrad 2009: 261), although the evidence is less robust than in other varieties.

5.7 Jamaican English (phase 4)

Although Jamaica was one of the most important colonies of the British Empire, trademark characteristics determining its linguistic development are the unique mixture of settler and indigenous groups (e.g. West African slaves, maroons, indentured laborers, sailors, convicts, farmers, planters, etc.; Schneider 2007: 228) and, at a later point, its geographical proximity to the US and the resulting AmE influence. Initially a Spanish possession, Jamaica was conquered by Britain in 1655 and developed into a sugar colony that predominantly relied on slave labor. Its colonial status continued basically unaltered through the official abolition of slavery at the beginning of the 19th century, and ended only in 1962 with Jamaican independence.[258]

Against the backdrop of the mixed population, the linguistic situation in Jamaica is veritably complex. While English is still the official language, a Creole variety has been developing since the end of the 17th century. This Creole was stigmatized until recently but an attitude shift toward accepting Jamaican Creole as an identity marker carrying more and more overt prestige has been observed particularly since independence. This has resulted in a split linguistic situation today, where BrE persists as a kind of exonormative target variety, although in practice Creole (also referred to as "Patois" or "Patwa") is increasingly accepted in spoken and also many written domains and seen as a matter of national pride. Standard BrE, on the other hand, especially in pronunciation, is associated with elitism (Schneider 2007: 233–236). Naturally, situational and social factors also play a crucial role in the variation between more or less standard-like usage (Schneider 1983a: 73–74), so that we can best speak of a continuum between basilectal Creole and a variety of educated JamE, whose distinctive features include lexical items and a characteristic accent (Schneider 2007: 235; cf. Patrick 1999). With reference to the future of this variety of English in Jamaica, linguists have

[258] For the settlement history of the Caribbean and Jamaica see further Sand (2002: 81–83).

doubted the continuing status of JamE as distinct (standardizing) variety (Wächtler 1977: 108–109), mainly because of the growing acceptance of Patois, which will possibly supplant JamE at one point (Sand 2002: 87–88). As has been noted above, another force that has exerted influence on JamE, in particular since the Second World War, is AmE (Hickey 2004: 14).

Although the grammar of (esp. written) JamE has been described as part of an "unmarked grammatical common core" (Schneider 2007: 235), Patois has also occasionally left its mark on educated JamE speech. One relevant instance is the non-marking of SPst forms (Sand 1999: 120; Patrick 2004: 416) and past participles or invariant use of PrTs or a base form in perfect contexts, as illustrated in (1) to (4).

(1) Has that *change* you (ICE-JA s1a-082)
(2) so it just happen all of a sudden and we have just *stick* with it [...] (ICE-JA s1a-061)
(3) This shop is safe; it *is* here for approximately 25 years. (Shields 1989: 50)
(4) But there this other girl that was there and shes[259] Jamaican as well and she never *have* no accent (ICE-JA s1a-022)

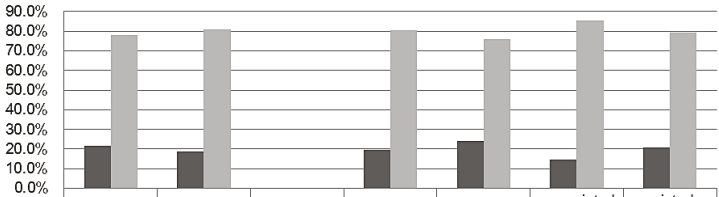
A related point is the statement that, comparable to PhiE, in both JamE and Patois an increased usage of adverbs (mostly *already* and *ever*) compensates for the non-marking of tense and aspect (Shields 1989: 50; cf. Trudgill & Hannah 2008: 115), although according to Patrick (1999: 185) this observation "proves to be inaccurate". Further features that are relevant for the perfect and which are potentially salient in this variety include, first, the absence of the auxiliary (Patrick 1999: 191; 2004: 416), which at times renders the distinction between PrPf and SPst forms impossible if no further clue is present in the context. Example (5) illustrates both points:

[259] At times, clitics (e.g. *shes, dont*; see also (5) and *its*, example (20) in Section 6.4 below) are transcribed like in this example in ICE-JA (spoken categories and w1b) and represented accordingly in the original data. However, since their absolute numbers are restricted (e.g. just one instance of *shes*), this did not affect the search for PrPfs.

(5) I dont know if you *ever heard* of the group P O K have uhm since uhm the election how skilfully they *ever fighted* and *denied* what the lyrics mean (ICE-JA s2a-043)

Second, the occasional usage of Creole tense and aspect markers such as *don(e)*, *ben*, preverbal *did* and *neva* has been mentioned (Patrick 1999: 171, 194–206; Durrleman-Tame 2008: 34–39; see further Section 6.4 and Deuber 2009: 26 on the role of *never/neva* in spoken JamE and Creole in particular). In addition, as regards the overall usage rates of the PrPf in JamE, earlier work has found a tendency to overuse the SPst in perfect contexts, while some speakers show the obverse tendency and overgeneralize the PrPf (Sand 1999: 119). The latter finding suggests that PrPfs in JamE may co-occur with definite temporal adverbials (see Section 6.6). For the present study, the findings of earlier work imply that JamE can be situated somewhere in the middle of the imagined hierarchy of PrPf-friendliness due to its relations to both BrE and AmE. The influence of Patois on JamE, too, limits the overall frequency of PrPfs that are used and potentially explains the occurrence of some non-standard features in the data, although these particularities will most likely be restricted to spoken text categories and will be comparatively rare overall. It remains to be tested whether marked differences exist between speech and writing with regard to the PrPf factors analyzed and whether the extended reliance on temporal adverbials that has been established by some observers for Patois can also be traced in JamE.

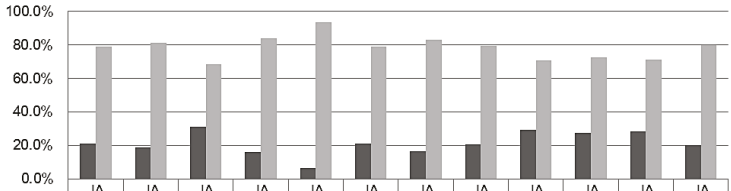
5.7.1 Temporal adverbials



	SPOKEN (n=331)	WRITTEN (n=211)		dialogues (n=173)	monologues (n=158)	non-printed writing (n=69)	printed writing (n=142)
with TA	21.8%	18.9%		19.7%	24.1%	14.5%	20.9%
without TA	78.2%	81.1%		80.3%	75.9%	85.5%	79.1%

Figure 5.7.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-JA (register and macro-genre differences)

Figure 5.7.1 displays the relative frequencies of PrPfs that are and are not temporally specified by an adverbial. In total, approximately one in five tokens in the JamE data is specified and no significant differences are found between writing and speech (where specification occurs more frequently) on the basis of the absolute counts. Among the macro-genres, monologues, with nearly one quarter of all tokens, has the highest relative rate of temporal adverbials, while specification is least common in non-printed writing, where only one in seven PrPfs is specified. None of the genres differs significantly from the overall distribution, from the values for the two registers or register-internally, however, which suggests a global homogeneity of the ICE-JA data in this respect.



	JA S1A (n=66)	JA S1B (n=107)	JA S2A (n=83)	JA S2B (n=75)	JA W1A (n=31)	JA W1B (n=38)	JA W2A (n=18)	JA W2B (n=53)	JA W2C (n=34)	JA W2D (n=11)	JA W2E (n=21)	JA W2F (n=5)
with TA	21.2%	18.7%	31.3%	16.0%	6.5%	21.1%	16.7%	20.8%	29.4%	27.3%	28.6%	20.0%
without TA	78.8%	81.3%	68.7%	84.0%	93.5%	78.9%	83.3%	79.2%	70.6%	72.7%	71.4%	80.0%

Figure 5.7.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-JA (detailed text categories)

More dispersion of the values occurs in the fine-grained view according to text types, which is shown in Figure 5.7.2. Values vary between 6.45% and 31.33% and the extreme points are represented by student writing (w1a) and unscripted monologues (s2a). Apart from the finding for the latter, the present analysis does not provide any further evidence in support of the above hypothesis that high proportions of temporal specification should occur in (some of) the spoken text categories due to a potential influence of Jamaican Creole.

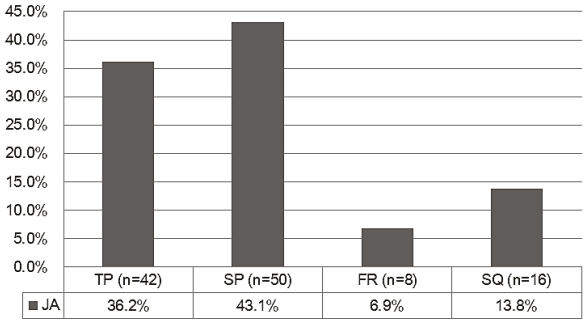


Figure 5.7.3. Types of PrPf adverbial specification: ICE-JA

The proportions of the types of temporal specification of PrPfs in ICE-JA are as follows (see Figure 5.7.3): Adverbials of span and duration (SP), as in (6), are most common and exceed two fifths of all PrPfs. Adverbials of time-position (TP), as in (7), with more than one third of all instances, are the second most frequent type. Example (8) illustrates adverbials of sequence (SQ), while example (9) illustrates frequency adverbials (FR), the type that appears only scarcely.

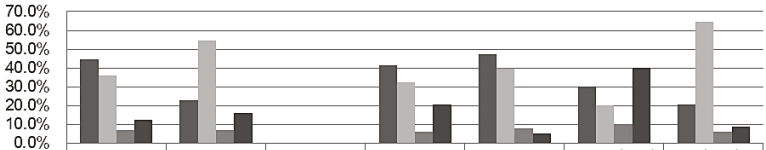
(6) Well watch what you are saying my boy because we have had something *from the nineteenth century* well documented probably the first to penetrate North America and so on (ICE-JA s2a-029)

(7) It is not only the leaders who have come *today* to pray (ICE-JA s2b-017)

(8) Because as we were talking earlier before uhm a a common practice now is that women who have acquired property *be-*

fore marriage dont want to accede that to the man [...] (ICE-JA s1b-033)

(9) You have called me here *several times* so how comes you dont have the number. (ICE-JA w1b-012)



	SPOKEN (n=72)	WRITTEN (n=44)		dialogues (n=34)	monologues (n=38)	non-printed writing (n=10)	printed writing (n=34)
TP	44.4%	22.7%		41.2%	47.4%	30.0%	20.6%
SP	36.1%	54.5%		32.4%	39.5%	20.0%	64.7%
FR	6.9%	6.8%		5.9%	7.9%	10.0%	5.9%
SQ	12.5%	15.9%		20.6%	5.3%	40.0%	8.8%

Figure 5.7.4. Types of PrPf adverbial specification: ICE-JA (register and macro-genre differences)

The relative frequencies of the various adverbial categories for the two registers (see Figure 5.7.4) indicate some marked differences: TP adverbials represent the most frequent type in speech, while SP adverbials cover more than half of all instances in writing. However, the variance in the relative values conceals the fact that differences between the two registers are not significant, as calculated using a Pearson Chi-square test. The distributions of the absolute values of the macro-genres also do not differ significantly from the overall or the respective register values. The spoken genres are internally consistent despite the large discrepancy in the relative SQ frequencies. In contrast, genre effects are present in the written genres, where a Fisher-exact test shows a significant difference between the two classes ($p < 0.05$). Note that this result has to be interpreted with caution due to the comparatively low token count for non-printed writing (10 tokens).

For the JamE data, the overall hierarchy of types of temporal specification that characteristically occur with the PrPf goes along the lines SP > TP > SQ > FR. We have seen, though, that differences in the modes of discourse and according to macro-genre exist. Like in PhiE, SQ adver-

bials cover a comparatively high share in dialogues, but are even more salient in non-printed writing. These adverbials could be used in a similar fashion in JamE, that is, as additional TR markers generalized from Patois usage. Given the scarcity of the data for the non-printed writing category, a follow-up study would be desirable to confirm this hypothesis. Nonetheless, in JamE, too, this phenomenon is likely to surface in the genre of dialogues, as the most involved and interactive category.

5.7.2 Aktionsart

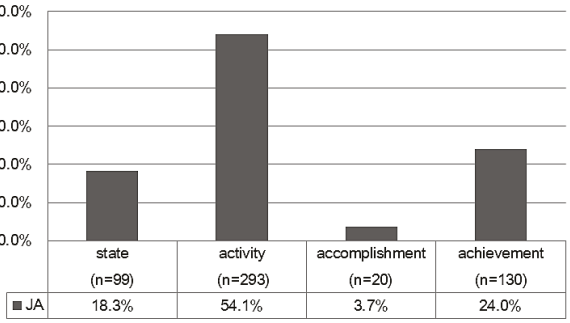
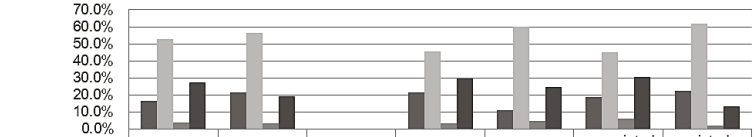


Figure 5.7.5. Aktionsart categories of main verbs in PrPf constructions: ICE-JA

The overall relative frequencies of Aktionsart categories of the JamE data are shown in Figure 5.7.5. More than half of all PrPf occurrences can be categorized as activity verbs (see (10)), while achievement verbs (see (11)) cover nearly one in four instances. Statives (see (12)) rank third and accomplishments (see (13)) constitute a marginal category in the ICE-JA sample.

(10) Apart from that, I have not *spoken* to any other person in Jamaica. (ICE-JA w1b-010)
(11) Will reparations solve the problem of mental slavery that many believe has *trapped* black people within the cycle of shame and guilt (ICE-JA s2b-025)
(12) Even some Orchid Society members have *been* at the forefront of these export drives. (ICE-JA w2b-026)

(13) The Caribbean since the twentieth century far from being a a a mere regurgitator of of European styles has *become* a major a consistent major exporter of culture especially in the area of music (ICE-JA s2a-029)



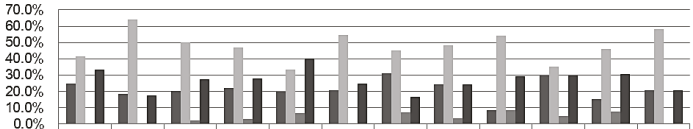
	SPOKEN (n=331)	WRITTEN (n=211)		dialogues (n=173)	monologues (n=158)	non-printed writing (n=69)	printed writing (n=142)
state	16.3%	21.3%		21.4%	10.8%	18.8%	22.5%
activity	52.6%	56.4%		45.7%	60.1%	44.9%	62.0%
accomplishment	3.9%	3.3%		3.5%	4.4%	5.8%	2.1%
achievement	27.2%	19.0%		29.5%	24.7%	30.4%	13.4%

Figure 5.7.6. Aktionsart categories of main verbs in PrPf constructions: ICE-JA (register and macro-genre differences)

Figure 5.7.6 displays the relative distribution of Aktionsart in PrPfs in ICE-JA, which appears to be fairly similar in the two registers; a Pearson Chi-square test confirms that the differences are not significant. Activity verbs are more common in both monologues and printed writing and this renders these two macro-genres significantly different from dialogues ($\chi^2 = 9.8962$, df = 3, $p < 0.05$, $\varphi_c = 0.173$) and non-printed writing ($\chi^2 = 11.714$, df = 3, $p < 0.01$, $\varphi_c = 0.236$) respectively. Printed writing is also the only macro-genre in which the distribution of Aktionsart values differs significantly from the overall distribution ($\chi^2 = 8.9776$, df = 3, $p < 0.05$, $\varphi_c = 0.115$), although with a weaker effect. Thus, we can conclude that some genre effects are present in ICE-JA in the area of Aktionsart of PrPf main verbs.

From the detailed display of Aktionsart proportions of the JamE data, which is presented in Figure 5.7.7, it transpires that activity verbs are the most frequent class in all text types apart from student writing (w1a). In the latter type, achievements are prominent, a finding which also applies to a number of further types, such as private dialogues (s1a), instructional (w2d) and persuasive (w2e) writing. A clear split between spoken ($\sigma = 0.04$) and written ($\sigma = 0.09$) text types in respect of the con-

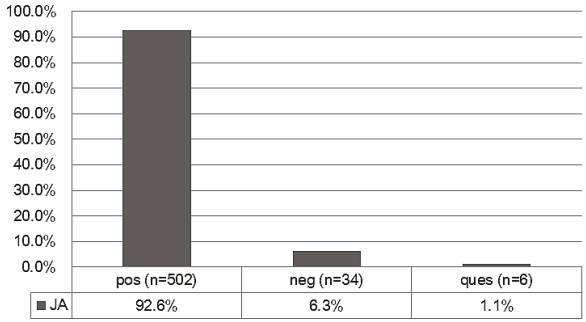
sistency of distributions can also be found in the ICE-JA sample, comparable to a number of other varieties particularly in speech.



	JA S1A (n=66)	JA S1B (n=107)	JA S2A (n=83)	JA S2B (n=75)	JA W1A (n=31)	JA W1B (n=38)	JA W2A (n=18)	JA W2B (n=53)	JA W2C (n=34)	JA W2D (n=11)	JA W2E (n=21)	JA W2F (n=5)
state	25.0%	18.4%	20.2%	21.9%	20.0%	20.6%	31.0%	24.1%	8.3%	30.0%	15.4%	20.8%
activity	41.7%	64.0%	50.0%	46.9%	33.3%	54.4%	45.2%	48.3%	54.2%	35.0%	46.2%	58.3%
accomplishment	0.0%	0.0%	2.4%	3.1%	6.7%	0.0%	7.1%	3.4%	8.3%	5.0%	7.7%	0.0%
achievement	33.3%	17.5%	27.4%	28.1%	40.0%	25.0%	16.7%	24.1%	29.2%	30.0%	30.8%	20.8%

Figure 5.7.7. Aktionsart categories of main verbs in PrPf constructions: ICE-JA (detailed text categories)

5.7.3 Sentence type



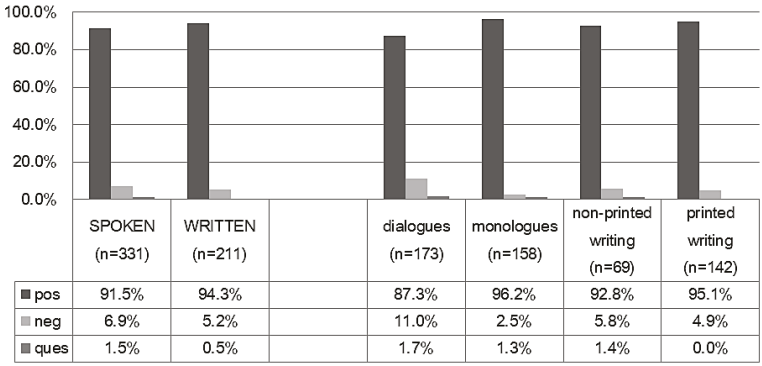
	pos (n=502)	neg (n=34)	ques (n=6)
JA	92.6%	6.3%	1.1%

Figure 5.7.8. Distribution of sentence types: ICE-JA

Figure 5.7.8 shows that in the ICE-JA sample the proportion of PrPfs that occur in positive statements exceeds nine out of ten tokens. Negatives, as in (14), surface in approximately every sixteenth token and questions, as exemplified in (15), are again marginal.

(14) I have it on good authority although I have not been able to locate the documentary evidence that Sir Philip later revised [...] (ICE-JA s2b-022)

(15) Do you know why Clarendon College has withdrawn from the Da Costa Cup (ICE-JA s2b-016)



	SPOKEN (n=331)	WRITTEN (n=211)		dialogues (n=173)	monologues (n=158)	non-printed writing (n=69)	printed writing (n=142)
pos	91.5%	94.3%		87.3%	96.2%	92.8%	95.1%
neg	6.9%	5.2%		11.0%	2.5%	5.8%	4.9%
ques	1.5%	0.5%		1.7%	1.3%	1.4%	0.0%

Figure 5.7.9. Distribution of sentence types: ICE-JA (register and macro-genre differences)

It is evident from Figure 5.7.9 that negatives are slightly more frequent in the spoken parts of the ICE-JA data. However, differences in the distributions of sentence types between the two modes of discourse do not reach the 5% level of statistical significance. As in a number of further varieties, the macro-genre of dialogues has the highest share of negatives, exceeding one in ten tokens. This renders the absolute distribution of sentence types in dialogues significantly different from monologues ($\chi^2 = 9.3253$, df = 2, $p < 0.01$, $\varphi_c = 0.168$). Otherwise, the distributions of the proportions of the macro-genres are similar (i.e. differences are not significant) both when the overall distribution and the values for the registers are taken as a baseline.

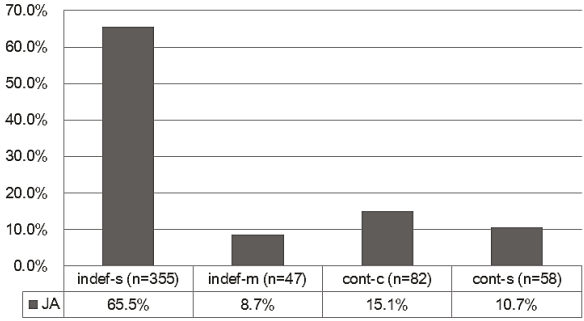
Figure 5.7.10 below shows that, given the overall rarity of the PrPf, it appears in questions only in a restricted group of text types. These are all spoken categories (apart from private dialogues (s1a)) and letters (w1b), arguably the most involved written category. While they are completely absent from various text types in the present sample, negatives are salient in dialogues (s1a and s1b) and furthermore in letters (w1b), reportage (w2c) and particularly in instructional writing (w2d). The finding for the last category has to be taken with caution as the overall number of tokens (11) is comparatively low. However, for the

remaining text types and dialogues in particular, results are based on larger token counts (> 30) and can be considered more robust. Therefore, it is evident that in the JamE data, text type (besides some minor genre influence) determines the distributions of the sentence types where PrPfs occur.

	JA S1A (n=66)	JA S1B (n=107)	JA S2A (n=83)	JA S2B (n=75)	JA W1A (n=31)	JA W1B (n=38)	JA W2A (n=18)	JA W2B (n=53)	JA W2C (n=34)	JA W2D (n=11)	JA W2E (n=21)	JA W2F (n=5)
pos	87.9%	86.9%	96.4%	96.0%	100.0%	86.8%	100.0%	100.0%	88.2%	72.7%	100.0%	100.0%
neg	12.1%	10.3%	2.4%	2.7%	0.0%	10.5%	0.0%	0.0%	11.8%	27.3%	0.0%	0.0%
ques	0.0%	2.8%	1.2%	1.3%	0.0%	2.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Figure 5.7.10. Distribution of sentence types: ICE-JA (detailed text categories)

5.7.4 Semantics



	indef-s (n=355)	indef-m (n=47)	cont-c (n=82)	cont-s (n=58)
JA	65.5%	8.7%	15.1%	10.7%

Figure 5.7.11. Distribution of semantic categories: ICE-JA

Figure 5.7.11 presents the breakdown of semantic categories of PrPfs in the JamE data. It emerges that indefinite single act/event readings (indef-s), illustrated in (16), cover nearly two thirds of all instances. Continuatives follow with act/event readings (cont-c), as in example (17), and states (cont-s), as in example (18), exceeding one in seven and one in ten

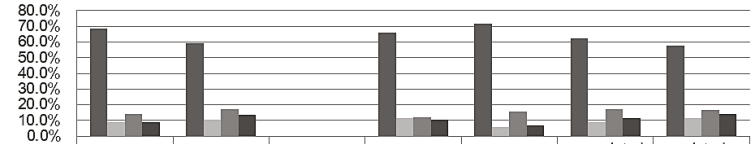
occurrences respectively. Indefinite multiple acts/events (indef-m), as illustrated in (19), are slightly less common than the latter class.

(16) The web has given any user the chance to express their individuality and ideals by creating personal web-sites. (ICE-JA w1a-020)

(17) [...] several eminent economists have sought to draw conclusions from the Balance of Trade figures and have expressed concern as to the meaning of these figures for the economy. (ICE-JA w2c-008)

(18) [...] everyone in a group in Surinam has preserved his language (ICE-JA s1b-011)

(19) In this somewhat bohemian setting, the gallery has held some ten exhibitions, showing such artists as Beti Campbell, Petrona Morrison, Mike Stanley and Giselle Valdez, as well as Colin himself. (ICE-JA w2b-003)

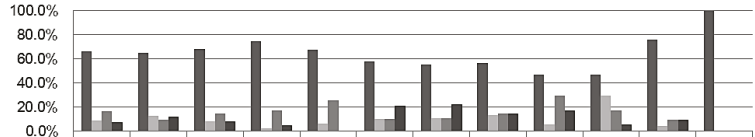


	SPOKEN (n=331)	WRITTEN (n=211)		dialogues (n=173)	monologues (n=158)	non-printed writing (n=69)	printed writing (n=142)
indef-s	68.6%	59.0%		65.9%	71.5%	62.3%	57.4%
indef-m	8.8%	10.6%		11.6%	5.7%	8.7%	11.5%
cont-c	13.9%	17.1%		12.1%	15.8%	17.4%	16.9%
cont-s	8.8%	13.4%		10.4%	7.0%	11.6%	14.2%

Figure 5.7.12. Distribution of semantic categories: ICE-JA (register and macro-genre differences)

The percentages of the semantic classes, as displayed in Figure 5.7.12, indicate some differences between the two registers, for instance higher shares for all classes apart from indef-s in writing. These differences are not significant, however. The comparison of the macro-genres both with each other and with the register and overall values also yields non-significant results, which suggests relative homogeneity of the distributions of the semantic categories in ICE-JA. Monologues, with a high

percentage of indef-s readings, is the only outlier, with a pattern significantly different from the overall distribution (χ^2 = 9.0651, df = 3, $p < 0.05$, $\varphi_C = 0.111$), while the effect size is comparatively small.



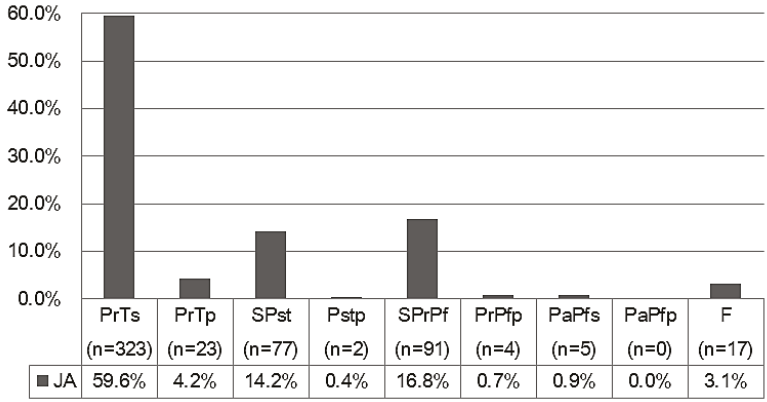
	JA S1A (n=66)	JA S1B (n=107)	JA S2A (n=83)	JA S2B (n=75)	JA W1A (n=31)	JA W1B (n=38)	JA W2A (n=18)	JA W2B (n=53)	JA W2C (n=34)	JA W2D (n=11)	JA W2E (n=21)	JA W2F (n=5)
indef-s	66.7%	65.4%	68.7%	74.7%	67.7%	57.9%	55.6%	56.6%	47.1%	47.1%	76.2%	100.0%
indef-m	9.1%	13.1%	8.4%	2.7%	6.5%	10.5%	11.1%	13.2%	5.9%	29.4%	4.8%	0.0%
cont-c	16.7%	9.3%	14.5%	17.3%	25.8%	10.5%	11.1%	15.1%	29.4%	17.6%	9.5%	0.0%
cont-s	7.6%	12.1%	8.4%	5.3%	0.0%	21.1%	22.2%	15.1%	17.6%	5.9%	9.5%	0.0%

Figure 5.7.13. Distribution of semantic categories: ICE-JA (detailed text categories)

Indef-s readings clearly constitute the most frequent semantic reading across all text types (see Figure 5.7.13). Once again, the distribution of the semantic classes is fairly similar for the spoken text types ($\sigma = 0.03$), while the written types are more dispersed ($\sigma = 0.10$). Conspicuous types are creative writing (w2f), where only indef-s occurs, an oddity that might be ascribed to the low token count (5) for this type; high percentages of cont-s readings occur in letters (w1b), academic writing (w2a) and reportage (w2c). The latter category also has the highest percentage of cont-c, which is comparable to the high proportion of indef-m in instructional writing (w2d).

All things considered – apart from the exceptions occurring in the fine-grained text type view – the picture that emerges from the analysis of the distribution of semantic categories in ICE-JA is a fairly homogeneous one.

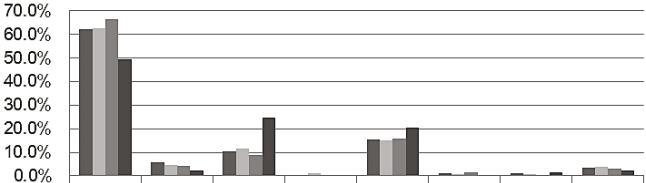
5.7.5 Preceding time reference



	PrTs (n=323)	PrTp (n=23)	SPst (n=77)	Pstp (n=2)	SPrPf (n=91)	PrPfp (n=4)	PaPfs (n=5)	PaPfp (n=0)	F (n=17)
JA	59.6%	4.2%	14.2%	0.4%	16.8%	0.7%	0.9%	0.0%	3.1%

Figure 5.7.14. TR forms preceding the PrPf: ICE-JA

The relative frequencies of preceding TR forms, as shown in Figure 5.7.14, bear no surprises: PrTs is the most frequent form, covering three fifths of all tokens. SPrPfs occur before more than every sixth and SPst before nearly every seventh PrPf. Further forms apart from PrTp and F are again negligible.



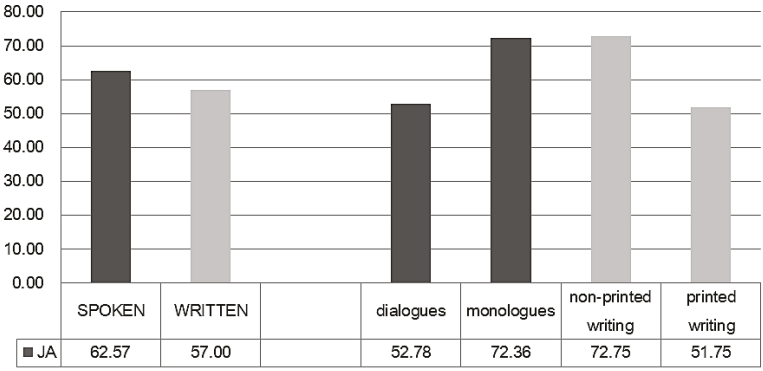
	PrTs (n=323)	PrTp (n=23)	SPst (n=77)	Pstp (n=2)	SPrPf (n=91)	PrPfp (n=4)	PaPfs (n=5)	F (n=17)
dialogues	62.4%	5.8%	10.4%	0.0%	15.6%	1.2%	1.2%	3.5%
monologues	62.7%	4.4%	11.4%	1.3%	15.2%	0.6%	0.6%	3.8%
non-printed writing	66.7%	4.3%	8.7%	0.0%	15.9%	1.4%	0.0%	2.9%
printed writing	49.3%	2.1%	24.6%	0.0%	20.4%	0.0%	1.4%	2.1%

Figure 5.7.15. TR forms preceding the PrPf: ICE-JA (macro-genre differences)

Figure 5.7.15 shows the macro-genre view. Here, as could be expected, PrTs, SPst and SPrPf are the only TR forms that are particularly salient.

Differences between the two registers (relative values not displayed in Figure 5.7.15) are not significant. However, while the proportions of the individual TR classes are fairly homogeneous for the remaining macro-genres, printed writing seems to be an outlier. Indeed, the printed writing genre is the only one whose differences border or exceed the 5% level of statistical significance when compared to both the overall distribution, as determined by a Pearson Chi-square test ($\chi^2 = 13.994$, df = 7, $p = 0.051$, $\varphi_C = 0.143$), and register-internally, as determined by a Fisher-exact test ($p < 0.05$), which was used due to a number of empty cells or cells with low values.

5.7.6 Perfect-friendliness and text type effects



	SPOKEN	WRITTEN		dialogues	monologues	non-printed writing	printed writing
JA	62.57	57.00		52.78	72.36	72.75	51.75

Figure 5.7.16. Frequency of PrPf occurrences in ICE-JA registers and macro-genres (normalized frequency per 10,000 words)

Based on the ICE data, the PrPf-friendliness score for JamE amounts to approximately 59 PrPfs/pttw and is thus comparatively low (third lowest out of ten varieties). This may be due to American or Creole influence, as explained above (or a combination of both). Figure 5.7.16 shows that PrPfs are more frequent in speech. When the macro-genres are considered, it transpires that differences between dialogues (53) and monologues (72) as well as between non-printed (73) and printed (52) writing are marked and cover roughly the same range.

The dispersion of the PrPf frequencies across the text types is moderate (σ = 24.19) and the extreme ends of the scale are represented by creative (w2f; 19) and instructional (w2d; 29) writing in opposition to scripted monologues (s2b; 83) and persuasive writing (w2e; 98). The values of the normalized PrPf frequencies correlate most strongly with those from IrE (r = 0.90), another variety that is situated at the lower end of the PrPf-friendliness scale. However, correlations with BrE (r = 0.85) and IndE (r = 0.82) are also strong.

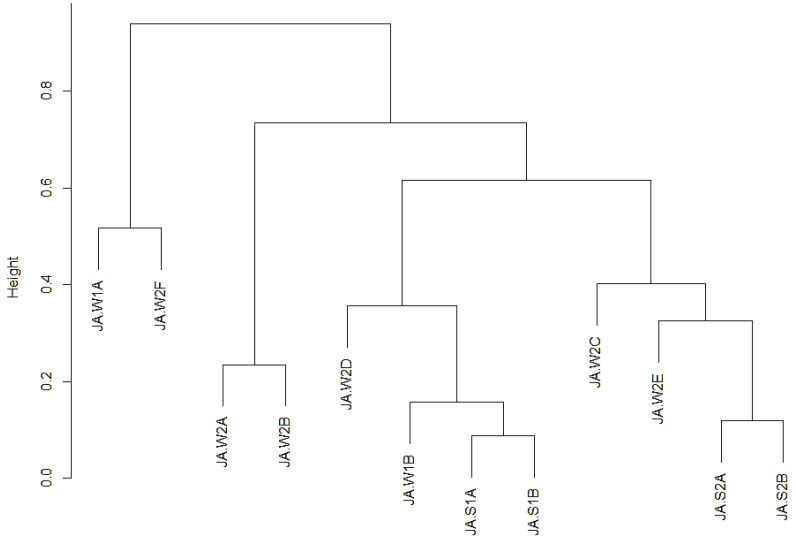


Figure 5.7.17. Cluster dendrogram of similarity across ICE text categories in JamE

The hierarchical cluster dendrogram (c = 0.80) for the JamE dataset, as shown in Figure 5.7.17, provides some evidence for alignment of spoken text types. Both the dialogues (s1a and s1b) and the monologues text types (s2a and s2b) can be found at directly neighboring leaves, while the two groupings appear to be clearly apart from each other. Letters (w1b) cluster with the dialogue text types at the next higher level.

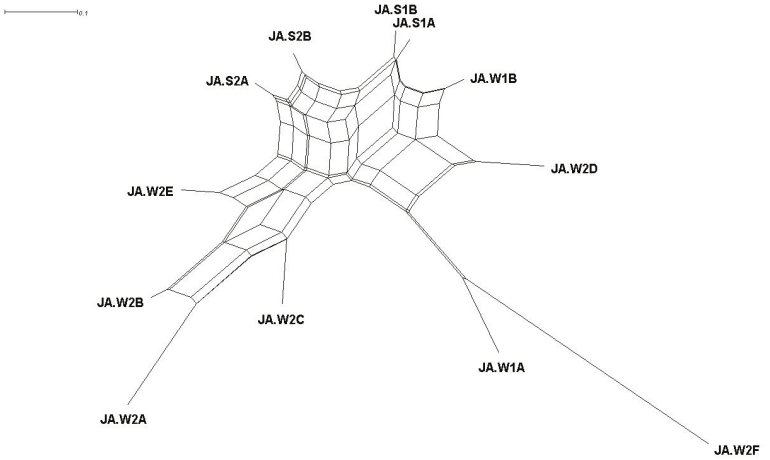


Figure 5.7.18. NeighborNet of similarity across ICE text categories in JamE

However, the non-hierarchical graphical representation in Figure 5.7.18 indicates that, in fact, all spoken text types can be found in adjacent nodes toward the top of the NeighborNet. A post-hoc multiscale bootstrap resampling test for significant clusters identifies the spoken triplet s1b, s2a, s2b along with w2e as significant (not shown above), but as the same applies to a written triplet (w2a, w2b, w2c), the results are somewhat inconclusive here. The closest written text type is that of letters (w1b), another textual class that is seen as involved, informal and comparable to dialogue. In contrast, the informational texts creative (w2f) and academic (w2a) writing, along with student writing (w1a), an involved category, appear distant, while the remaining text types cover intermediate positions. Yet again, the JamE data on the PrPf, too, can be interpreted as evidence for the closeness or (structural) uniformity of spoken texts *sensu* Biber & Conrad (2009: 261).

5.8 Singaporean English (phase 4)

SinE represents one of the urban Southeast Asian varieties of English. The Singaporean population is commonly characterized as a blend of various strands (predominantly people of Chinese, Indian, Malay and also Arab and European descent) resulting in a multiethnic and multi-

lingual society (see Vaish 2007 or Ansaldo 2010: 502–505 for sociolinguistic detail). Singapore was established in 1819 as a strategic outpost of the East India Company – a direct outcome of British colonial endeavors – and subsequently developed into an important center of commerce. In 1867 Singapore became a Crown colony, but voices for independence became louder over time, in particular after a short period of Japanese occupation from 1942 to 1945 and the end of the Second World War. Periods of self-government (1959–1963) and the establishment as independent city-state (1963) followed in the 1960s, interrupted by a short-lived unification with Malaysia from 1963 to 1965 (Schneider 2007: 153–155).

Due to its colonial past, the present state of affairs in Singapore with regard to culture in general and the linguistic situation in particular has been characterized as "hybrid" (Schneider 2011: 158) in multiple respects. Above all, early dialect contact and mixing among the English-speaking settler population occurred, involving Scots, Irish and American as well as British varieties (Schneider 2007: 154–155; Davydova 2013). Second, various substrates such as different varieties of Chinese, Malay and Indian languages have impacted the linguistic configuration. Yet, English has been viewed as the language that is most beneficial for the international economic competitiveness of the country (Kachru & Nelson 2006: 184). At the same time, it is considered neutral and exerts an integrating force across the various population groups of diverse ethnic descent. In addition, while numbers of English-educated people in Singapore traditionally had been high compared to other Asian countries (Brutt-Griffler 2002: 94), since 1987 the government has endorsed a strong language policy of bilingualism. English is the compulsory first language and one of the officially recognized ethnic languages (either standard Mandarin Chinese, Malay or Tamil) is taught as a second language, irrespective of the varieties of these languages spoken at home and the ensuing social implications (Schneider 2007: 156). Besides a clear dominance of English in public environments, this policy has in the long run led to a marked increase of English also in personal domains, a development which seems to be continuing among younger generations of Singaporeans. In practice, however, the English of Singapore by no means represents a uniform variety (Tickoo 1996: 446) and

varies according to social and situational parameters. The most conspicuous variant is the basilectal form referred to as "Singlish". It is structurally characterized by influence from the Chinese and Malay substrates [260] and carries notions of informality and emotionality. Note that the issue of Singlish has been hotly debated in Singaporean society; while a recent governmental initiative rejects its recognition, it has gained covert prestige as a marker of local Singaporean identity and pride. Toward the acrolectal pole of the dialect continuum a standard variety of SinE, allegedly still showing traces of exonormative orientation toward BrE (Pennycook 1994: 235–236; Tickoo 1996: 447; Bautista 2008: 217; but cf. Schneider 2007: 160–161), can be found. It is mainly characterized by distinctive features in pronunciation and vocabulary, while the mesolect is naturally present in many linguistic shapes (Schneider 2007: 157–160; on Singlish see further Schneider 2011: 159–160).

The grammar of SinE has also been described as "hybrid" (Ansaldo 2004: 136). Traces of the substrates (potentially via Singlish) can be found in colloquial SinE, although some have argued that substrate-derived and English-derived grammatical systems are clearly kept separate according to formality of the situation (Bao 2005: 264–265; but cf. Platt 1982: 402). In a similar fashion, it has also been argued that a marked difference emerges according to mode of discourse. While writing is strongly oriented toward (educated) standard usage, local usage is prominent in spoken registers (Pennycook 1994: 235–236). Naturally, these dimensions (mode of discourse/formality) correlate to a certain degree. As in other L2 varieties of English (e.g. JamE or HKE), speakers of SinE sporadically tend to use PrTs or base forms once a time frame has been established (Lian 2003: 46–47; Gut 2009: 273–274). [261] As an alternative way of expressing time reference, temporal adverbials, in particular *ever* and *already*, as in examples (1) to (4) below, [262] have been identified as grammaticalized markers of the perfect that are in use instead of a TR form (Wee 2004: 1058–1059; Deterding 2008: 52; Sharma

[260] For an in-depth analysis of the various substrates of SinE see Sharma (2009: 175).
[261] Davydova mentions that an extended functional range of the PrTs may also be partly attributable to the influence of an IrE superstrate of the 19th century (2011: 255).
[262] Note also the Singlish discourse markers *lah* in (3) and *know* in (4).

2009: 179; Szmrecsanyi 2012; see also sections 2.1 and 6.5),[263] or in combination with a SPst (Platt 1982: 399):

(1) Maybe she increase the price *already* (ICE-SIN s1a-006)
(2) Uhm I yearn for one week *already* (ICE-SIN s1a-013)
(3) Actually we challenge him for two years *already* lah (ICE-SIN s1a-013)
(4) I *ever* bought a jacket a sweater know two dollars (ICE-SIN s1a-057)

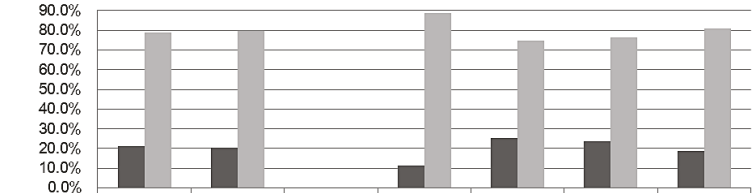
The findings from earlier work can be applied to the present study insofar as the use of PrTs or base forms instead of the PrPf may result in a lower degree of PrPf-friendliness, particularly in informal speech. Another line of argumentation that would support a low PrPf-friendliness is a potential re-orientation toward AmE due to the strong influence American pop and media culture has exerted on Singaporeans (Köppl 1983b: 123; Pennycook 1994: 223). Furthermore, it remains to be established whether speech and writing differ with regard to the PrPf factors analyzed and whether the extended reliance on temporal adverbials as additional markers of perfectivity is also in evidence in the ICE-SIN data.

5.8.1 Temporal adverbials

The proportions of PrPfs with and without specification by a temporal adverbial are shown in Figure 5.8.1. Around one fifth of all occurrences are specified and differences between the spoken and the written part are not significant. The extreme ends of the spectrum are to be found within the spoken macro-genres: monologues, with more than one in four instances, has the highest, while dialogues, with one in nine only, has the lowest relative value for specification, which results in significant differences between these macro-genres ($\chi^2 = 5.3855$, $df < 1$, $p = 0.05$, $\varphi_c = 0.159$). However, none of the distributions of the genres is significantly different from the overall distribution and, apart from dialogues

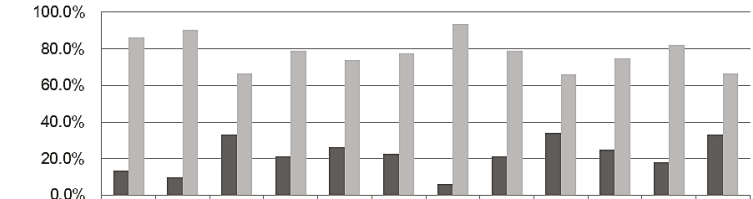
[263] This goes hand in hand with a tendency to mark aspectual rather than tense distinctions, as these are "robust features of Sinitic and Malay alike" (Ansaldo 2010: 509).

($\chi^2 = 5.0930$, df = 1, $p < 0.05$, $\varphi_c = 0.134$), from the respective aggregated register distributions. All in all, this represents a mixed result for the SinE data as regards homogeneity.



	SPOKEN (n=213)	WRITTEN (n=264)		dialogues (n=63)	monologues (n=150)	non-printed writing (n=68)	printed writing (n=196)
with TA	21.1%	20.3%		11.1%	25.3%	23.5%	18.9%
without TA	78.9%	79.7%		88.9%	74.7%	76.5%	81.1%

Figure 5.8.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-SIN (register and macro-genre differences)



	SIN S1A (n=22)	SIN S1B (n=41)	SIN S2A (n=51)	SIN S2B (n=99)	SIN W1A (n=19)	SIN W1B (n=49)	SIN W2A (n=32)	SIN W2B (n=38)	SIN W2C (n=44)	SIN W2D (n=28)	SIN W2E (n=39)	SIN W2F (n=15)
with TA	13.6%	9.8%	33.3%	21.2%	26.3%	22.4%	6.3%	21.1%	34.1%	25.0%	17.9%	33.3%
without TA	86.4%	90.2%	66.7%	78.8%	73.7%	77.6%	93.8%	78.9%	65.9%	75.0%	82.1%	66.7%

Figure 5.8.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-SIN (detailed text categories)

Naturally, values are more dispersed in the detailed text category view, as displayed in Figure 5.8.2. Academic writing (w2a; 6.3%) and news reportage (w2c; 34.1%) cover the extreme poles. It is further revealed that, apart from the result for unscripted monologues (s2a), the data do not support the hypothesis formulated above that a high rate of temporal specification should be evident particularly in the spoken text categories (see further Section 6.4). In fact, the dialogue text categories yield speci-

fication values that are well below the overall ICE-SIN average and the involved written categories also do not yield an extensive proportion of specification.

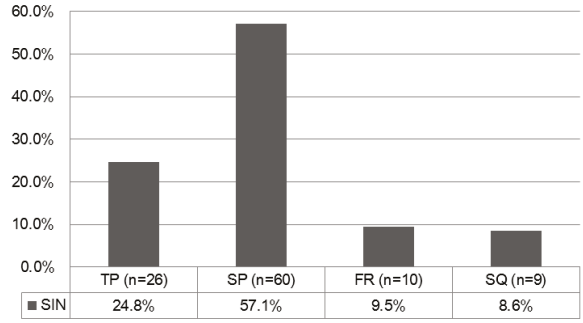


Figure 5.8.3. Types of PrPf adverbial specification: ICE-SIN

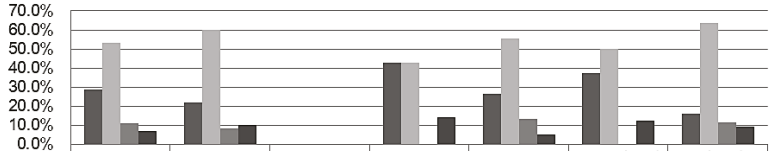
Figure 5.8.3 shows the distribution of the classes of adverbial specification in the SinE data. Again, adverbials of span and duration (SP), as in example (5), represent the most frequent category, with nearly three fifths of all tokens. The second most common class is adverbials of time-position (TP), as in (6), with nearly one quarter of all instances. The remaining two classes, frequency adverbials (FR), as in (7), and adverbials of sequence (SQ), as in example (8), cover roughly equal shares, with fewer than one out of ten tokens each.

(5) [...] the fact that the Singapore market has been able to hold up so very well *during this period* given the very close correlation between these two indices I think it's indicative of the fact that there is strong buying out there (ICE-SIN s1b-028)

(6) Japan's parliament has resumed debate *after a two week stalemate* (ICE-SIN s2b-013)

(7) this is *the fifteen time* that Singapore has won and they have not lost one single game (ICE-SIN s2a-018)

(8) But *so far* it's been sunny with bits and pieces of rain. (ICE-SIN w1b-010)



	SPOKEN (n=45)	WRITTEN (n=60)		dialogues (n=7)	monologues (n=38)	non-printed writing (n=16)	printed writing (n=44)
TP	28.9%	21.7%		42.9%	26.3%	37.5%	15.9%
SP	53.3%	60.0%		42.9%	55.3%	50.0%	63.6%
FR	11.1%	8.3%		0.0%	13.2%	0.0%	11.4%
SQ	6.7%	10.0%		14.3%	5.3%	12.5%	9.1%

Figure 5.8.4. Types of PrPf adverbial specification: ICE-SIN (register and macro-genre differences)

Figure 5.8.4 displays the distributions of the adverbial classes in the two modes of discourse. It transpires that SP adverbials are most common in both registers, while TP and FR are slightly less and SQ adverbials slightly more prominent in writing. However, a Pearson Chi-square test reveals that the differences are not significant. The same applies to the comparisons of the breakdown for each macro-genre with the overall values, the respective register values and also register-internally, although no FR adverbials are present in both dialogues and non-printed writing. Yet, on the basis of the above statistical tests we can assume that the distributions of adverbial classes are fairly uniform across the genres.

The hierarchy of classes of temporal adverbial specification co-occurring with the PrPf can be established in the ICE-SIN data as follows: SP > TP > FR > SQ. Similar to PhiE and JamE, SQ adverbials, that is, those that have been identified as salient TR markers (esp. *already* and *ever*), are considerably more frequent in dialogues, a finding that may hint at Singlish-derived usage patterns partly traceable in mesolectal and acrolectal speech. While more data would be desirable to confirm or reject this hypothesis (see further sections 6.4 and 6.5), this phenomenon is probably most likely to occur in dialogues, as the most involved and interactive ICE textual category.

5.8.2 Aktionsart

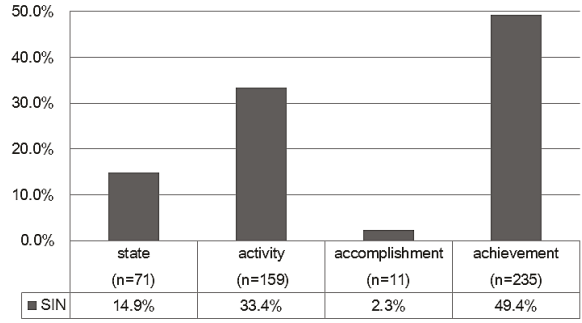


Figure 5.8.5. Aktionsart categories of main verbs in PrPf constructions: ICE-SIN

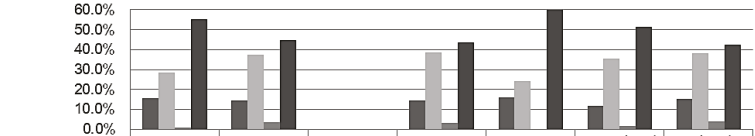
Figure 5.8.5 presents the breakdown of the Aktionsart categories. In the ICE-SIN data, nearly half of all PrPfs can be classed as achievements, as exemplified in (9), and one third of all PrPfs as activities, as exemplified in (10). Statives, as in (11), cover approximately one in seven instances and accomplishments, as in (12), are rare.

(9) [...] I had better write you something before you're again "SOMEWHERE OUT THERE" and where nobody can reach you by mail/air/phone or whatever the latest technology has *invented*. (ICE-SIN w1b-008)

(10) She has *lived* through all the wars and her son Ding Fong's brother was killed in Dian Bian Fu (ICE-SIN s2b-021)

(11) An emphasis on miniatures has not hitherto *been* a feature of tropical growing, but there is no reason why it should not be, as the number of would-be growers with restricted space increases. (ICE-SIN w2d-016)

(12) Ya but it is like transported because of some Western values you see so the thing is whether that problem uh has *become* a international thing (ICE-SIN s1a-065)

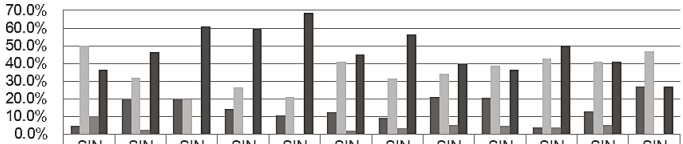


	SPOKEN (n=213)	WRITTEN (n=264)		dialogues (n=63)	monologues (n=150)	non-printed writing (n=68)	printed writing (n=196)
state	15.6%	14.4%		14.5%	16.0%	11.8%	15.3%
activity	28.3%	37.5%		38.7%	24.0%	35.3%	38.3%
accomplishment	0.9%	3.4%		3.2%	0.0%	1.5%	4.1%
achievement	55.2%	44.7%		43.5%	60.0%	51.5%	42.3%

Figure 5.8.6. Aktionsart categories of main verbs in PrPf constructions: ICE-SIN (register and macro-genre differences)

Figure 5.8.6, which shows the breakdown of the Aktionsart categories of PrPfs in the SinE data, yields differences between the two registers, with more activities but fewer achievements in writing. A Pearson Chi-square test shows that these differences are significant ($\chi^2 = 8.8013$, df = 3, $p < 0.05$, $\varphi_c = 0.136$). Achievement verbs are the most frequent Aktionsart category across all macro-genres, and particularly prominent in monologues, the only macro-genre that is significantly different from the overall breakdown ($\chi^2 = 10.406$, df = 3, $p < 0.05$, $\varphi_c = 0.222$). Significant register-internal differences also exist between monologues and dialogues ($\chi^2 = 9.2749$, df = 3, $p < 0.05$, $\varphi_c = 0.122$), while the two written genres are comparatively homogeneous when compared both to each other and to the overall distribution. These findings suggest that a number of register and genre effects exist in the SinE data as regards the distribution of the Aktionsart classes of PrPfs.

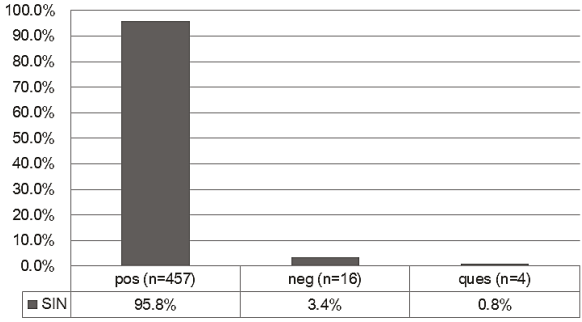
Figure 5.8.7, which displays the detailed proportions, illustrates the variability present in the ICE-SIN data. It is noticeable that within the rare category of accomplishments proportions vary between 0% and 9.5% and that achievements are not the most frequent class in a number of textual categories (private dialogues (s1a); reportage (w2c); creative writing (w2f)), where activities are much more prominent. Compared to other varieties (e.g. JamE), the ICE-SIN data do not show a clear split between the spoken ($\sigma = 0.06$) and written ($\sigma = 0.06$) text categories as regards the consistency of the distributions that was in evidence for the former register.



	SIN S1A (n=22)	SIN S1B (n=41)	SIN S2A (n=51)	SIN S2B (n=99)	SIN W1A (n=19)	SIN W1B (n=49)	SIN W2A (n=32)	SIN W2B (n=38)	SIN W2C (n=44)	SIN W2D (n=28)	SIN W2E (n=39)	SIN W2F (n=15)
state	4.5%	19.5%	19.6%	14.1%	10.5%	12.2%	9.4%	21.1%	20.5%	3.6%	12.8%	26.7%
activity	50.0%	31.7%	19.6%	26.3%	21.1%	40.8%	31.3%	34.2%	38.6%	42.9%	41.0%	46.7%
accomplishment	9.5%	2.4%	0.0%	0.0%	0.0%	2.0%	3.1%	5.3%	4.5%	3.6%	5.1%	0.0%
achievement	36.4%	46.3%	60.8%	59.6%	68.4%	44.9%	56.3%	39.5%	36.4%	50.0%	41.0%	26.7%

Figure 5.8.7. Aktionsart categories of main verbs in PrPf constructions: ICE-SIN (detailed text categories)

5.8.3 Sentence type



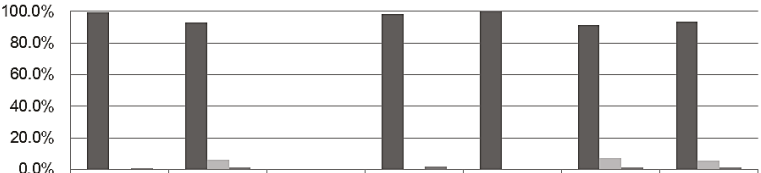
	pos (n=457)	neg (n=16)	ques (n=4)
SIN	95.8%	3.4%	0.8%

Figure 5.8.8. Distribution of sentence types: ICE-SIN

In the SinE data, almost all PrPf tokens occur in positive statements (see Figure 5.8.8), while negatives, as in example (11), repeated here as (13), are rare, and questions, as in example (14), are least common.

(13) An emphasis on miniatures has not hitherto been a feature of tropical growing, but there is no reason why it should not be, as the number of would-be growers with restricted space increases. (ICE-SIN w2d-016)

(14) What has happened to this guy (ICE-SIN s1a-087)



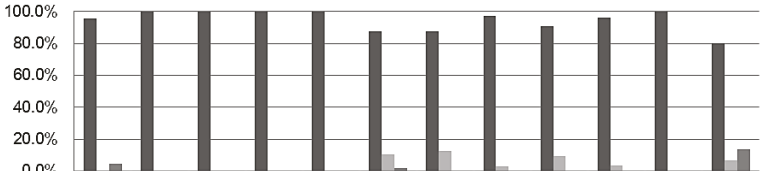
	SPOKEN (n=213)	WRITTEN (n=264)		dialogues (n=63)	monologues (n=150)	non-printed writing (n=68)	printed writing (n=196)
pos	99.5%	92.8%		98.4%	100.0%	91.2%	93.4%
neg	0.0%	6.1%		0.0%	0.0%	7.4%	5.6%
ques	0.5%	1.1%		1.6%	0.0%	1.5%	1.0%

Figure 5.8.9. Distribution of sentence types: ICE-SIN (register and macro-genre differences)

Figure 5.8.9 reveals that forms other than positives almost never occur (one single question) in the speech section of ICE-SIN, which results in highly significant differences when contrasted to writing ($\chi^2 = 14.901$, df = 2, $p < 0.001$, $\varphi_C = 0.172$). In light of the findings for other varieties, it comes as a surprise that no negatives are present in speech and in dialogues in particular. With regard to the internal consistency of the distributions of the sentence types, there are mixed results: While monologues is the only macro-genre that yields significant differences from the overall distribution, although with a weak effect ($\chi^2 = 6.4965$, df = 2, $p < 0.05$, $\varphi_C = 0.102$), no other statistical tests (i.e. when the values for the macro-genres are compared with the overall distribution, with the respective register distribution and register-internally by a Fisher exact test) yield significant results, which in turn suggests relative homogeneity of the data.

Figure 5.8.10 shows that questions appear in three text categories (private dialogues (s1a); letters (w1b); creative writing (w2f)) only. Note, however, that the relative values are slightly deceptive here as the samples for each text type are small, such that the relative values translate into one (s1a, w1b) or two (w2f) tokens only. Note that in the last text type dialogue or direct speech is narrated (see example (15)):

(15) "You have been friends with Julie a long time?" she asked. "Yes ...." I stammered, "I mean .... no ...." (ICE-SIN w2f-005)

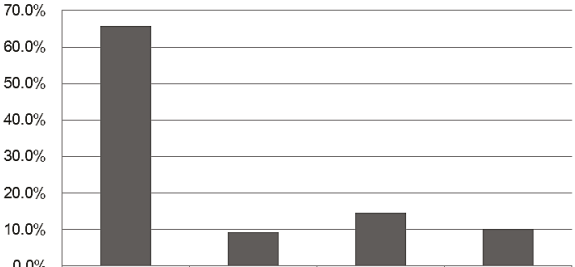


	SIN S1A (n=22)	SIN S1B (n=41)	SIN S2A (n=51)	SIN S2B (n=99)	SIN W1A (n=19)	SIN W1B (n=49)	SIN W2A (n=32)	SIN W2B (n=38)	SIN W2C (n=44)	SIN W2D (n=28)	SIN W2E (n=39)	SIN W2F (n=15)
pos	95.5%	100.0%	100.0%	100.0%	100.0%	87.8%	87.5%	97.4%	90.9%	96.4%	100.0%	80.0%
neg	0.0%	0.0%	0.0%	0.0%	0.0%	10.2%	12.5%	2.6%	9.1%	3.6%	0.0%	6.7%
ques	4.5%	0.0%	0.0%	0.0%	0.0%	2.0%	0.0%	0.0%	0.0%	0.0%	0.0%	13.3%

Figure 5.8.10. Distribution of sentence types: ICE-SIN (detailed text categories)

It is also worth noting that negatives are absent from student (w1a) and persuasive writing (w2e) as well as from all spoken types, as indicated above, while they are salient in letters (w1b) and academic writing (w2a). In sum, the ICE-SIN data suggest that text type influences the breakdown of sentence types in PrPf contexts, while effects on the more general levels (register and genre) are only minor.

5.8.4 Semantics



	indef-s (n=314)	indef-m (n=44)	cont-c (n=70)	cont-s (n=49)
SIN	65.8%	9.2%	14.7%	10.3%

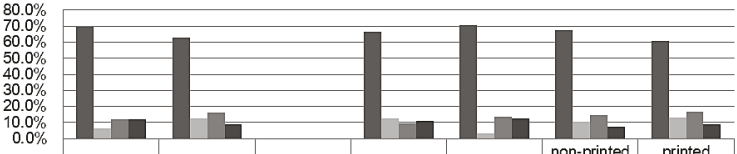
Figure 5.8.11. Distribution of semantic categories: ICE-SIN

The distribution of semantic categories with PrPf verbs in the ICE-SIN sample, as presented in Figure 5.8.11, illustrates that indefinite single acts/events (indef-s), as in (16), represent a highly frequent category with approximately two thirds of all tokens. Continuative acts/events (cont-c),

as exemplified in (17), cover more than every seventh token, while tokens with a continuative state (cont-s; see (18)) and indefinite multiple act reading (indef-m; see (19)) cover around one tenth of the data each.

(16) The Soviet Union has called for the immediate and unconditional withdrawal of the Iraqi forces. (ICE-SIN w2e-001)
(17) So far, five people have contracted Aids through intravenous drug abuse. (ICE-SIN w2c-013)
(18) Wet spring weather has left farmers stuck in the mud delaying the seeding of billions of dollars worth of crops (ICE-SIN s2b-020)
(19) As much as we know that Satan has sinned and is therefore in the wrong, he still emerges as a hero. (ICE-SIN w1a-001)

Some dissimilarity between the two modes of discourse occurs in the distribution of the semantic categories (see Figure 5.8.12), with a smaller share of indef-s and cont-s but higher proportions in the other two categories in writing.

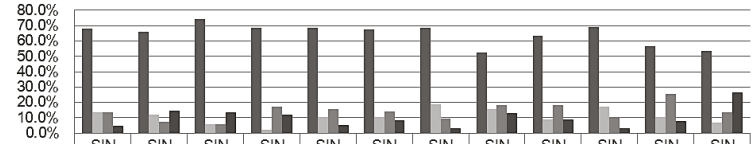


	SPOKEN (n=213)	WRITTEN (n=264)		dialogues (n=63)	monologues (n=150)	non-printed writing (n=68)	printed writing (n=196)
indef-s	69.5%	62.6%		66.7%	70.7%	67.6%	60.9%
indef-m	6.1%	12.5%		12.7%	3.3%	10.3%	13.2%
cont-c	12.2%	16.2%		9.5%	13.3%	14.7%	16.8%
cont-s	12.2%	8.7%		11.1%	12.7%	7.4%	9.1%

Figure 5.8.12. Distribution of semantic categories: ICE-SIN (register and macro-genre differences)

A Pearson Chi-square test reveals that these differences are statistically significant ($\chi^2 = 8.5438$, df = 3, $p < 0.05$, $\varphi_c = 0.134$). In contrast, although some variance can be observed in the relative values, none of the macro-genres is significantly different from the overall and the respective register breakdown, nor when compared register-internally. There-

fore, while some register effects are in evidence, we can establish a fairly uniform distribution of the semantic categories on the basis of the ICE-SIN dataset.

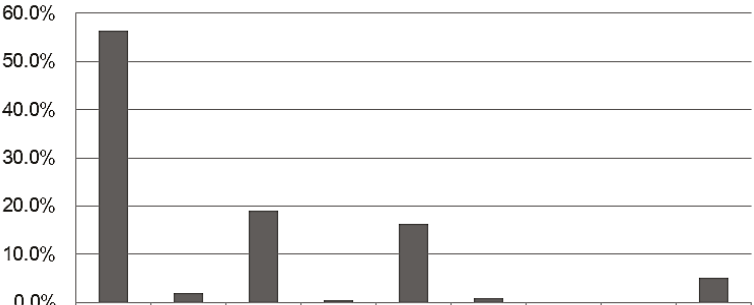


	SIN S1A (n=22)	SIN S1B (n=41)	SIN S2A (n=51)	SIN S2B (n=99)	SIN W1A (n=19)	SIN W1B (n=49)	SIN W2A (n=32)	SIN W2B (n=38)	SIN W2C (n=44)	SIN W2D (n=28)	SIN W2E (n=39)	SIN W2F (n=15)
indef-s	68.2%	65.9%	74.5%	68.7%	68.4%	67.3%	68.8%	52.6%	63.6%	69.0%	56.4%	53.3%
indef-m	13.6%	12.2%	5.9%	2.0%	10.5%	10.2%	18.8%	15.8%	9.1%	17.2%	10.3%	6.7%
cont-c	13.6%	7.3%	5.9%	17.2%	15.8%	14.3%	9.4%	18.4%	18.2%	10.3%	25.6%	13.3%
cont-s	4.5%	14.6%	13.7%	12.1%	5.3%	8.2%	3.1%	13.2%	9.1%	3.4%	7.7%	26.7%

Figure 5.8.13. Distribution of semantic categories: ICE-SIN (detailed text categories)

This finding is further confirmed by the fine-grained text category view, which is presented in Figure 5.8.13. While the variance in the spoken text types amounts to $\sigma = 0.04$, the value for the written texts ($\sigma = 0.06$) is comparatively low. Indef-s readings consistently constitute the most frequent semantic reading across the board, exceeding more than half of all PrPfs. Outliers in the less frequent semantic categories are scripted monologues (s2b), with only few indef-m readings, and instructional writing (w2d), with few instances of cont-s. The opposite is the case in creative writing (w2f), where more than one in four tokens has a cont-s reading; and the same applies to the share of cont-c in persuasive writing (w2e). Notwithstanding these peculiarities of the detailed analysis and the register effect identified above, the breakdown of semantic categories in the ICE-SIN data is homogeneous.

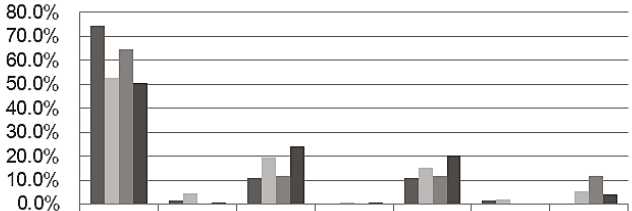
5.8.5 Preceding time reference



	PrTs (n=269)	PrTp (n=9)	SPst (n=91)	Pstp (n=2)	SPrPf (n=78)	PrPfp (n=4)	PaPfs (n=0)	PaPfp (n=0)	F (n=24)
SIN	56.4%	1.9%	19.1%	0.4%	16.4%	0.8%	0.0%	0.0%	5.0%

Figure 5.8.14. TR forms preceding the PrPf: ICE-SIN

As in a number of other varieties, only a restricted set of preceding TR forms appears in perceptible quantities (see Figure 5.8.14). PrTs again represents the majority variant, with well over half of all tokens, while SPst and SPrPf forms occur before every fifth and sixth token respectively. In addition, some PrPfs are preceded by F and PrTp forms, while further TR forms occur only sporadically.



	PrTs (n=269)	PrTp (n=9)	SPst (n=91)	Pstp (n=2)	SPrPf (n=78)	PrPfp (n=4)	F (n=24)
dialogues	74.6%	1.6%	11.1%	0.0%	11.1%	1.6%	0.0%
monologues	52.7%	4.7%	19.3%	0.7%	15.3%	2.0%	5.3%
non-printed writing	64.7%	0.0%	11.8%	0.0%	11.8%	0.0%	11.8%
printed writing	50.5%	0.5%	24.0%	0.5%	20.4%	0.0%	4.1%

Figure 5.8.15. TR forms preceding the PrPf: ICE-SIN (macro-genre differences)

The macro-genre view, presented in Figure 5.8.15, again shows that forms other than PrTs, SPst and SPrPf are rare. In the SinE data, speech and writing behave significantly differently (relative values not shown; χ^2= 16.037, df = 6, p < 0.05, φ_c = 0.183). This is mainly due to non-printed writing, where the shares of SPst, SPrPf and F are equal, which also renders this macro-genre significantly different from printed writing, as established by a Fisher exact test (p < 0.05) used due a number of empty cells or cells with low counts. Otherwise, the data are homogeneous; that is, none of the macro-genres yields proportions of the absolute values that are significantly different from the overall or register breakdown.

5.8.6 Perfect-friendliness and text type effects

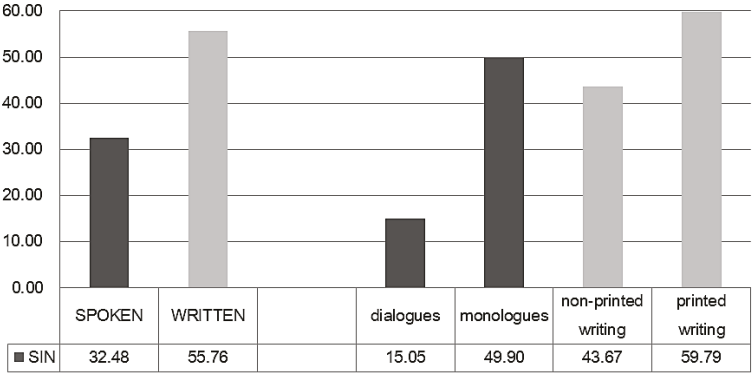


Figure 5.8.16. Frequency of PrPf occurrences in ICE-SIN registers and macro-genres (normalized frequency per 10,000 words)

The SinE sample, with 48 PrPfs/pttw overall, is the least PrPf-friendly in the whole dataset. As indicated above, this low value can be interpreted as evidence for a tendency toward Americanization combined with strong substrate influence (non-marking of TR forms). What is most striking in Figure 5.8.16 is the very low rate of PrPfs in dialogues (15), the most involved and interactive macro-genre, which contrasts strongly with monologues (50) and also translates into a considerably lower value for speech overall. This result is mainly attributable to private dialogues

(s1a), where PrPfs occur least frequently of all text types of all varieties included in the present analysis (see also below). The values for the written registers are also well below average, with ICE-SIN being the least PrPf-friendly variety in non-printed writing (44) and the third least PrPf-friendly in printed writing (60). Further, ICE-SIN is the only dataset, apart from ICE-CAN, where PrPf-friendliness is higher in writing than in speech.

At the same time, SinE has the second highest dispersion value of all varieties (σ = 32.19), with values ranging from eight (private dialogues, s1a) to 133 (persuasive writing, w2e). This result could be seen as illustrative of the existence of the situational internal variability of SinE as indicated above. The strongest correlation of the normalized PrPf frequencies across the detailed text types is in evidence with IrE (r = 0.76), a variety that can be found toward the lower end of the PrPf-friendliness scale.

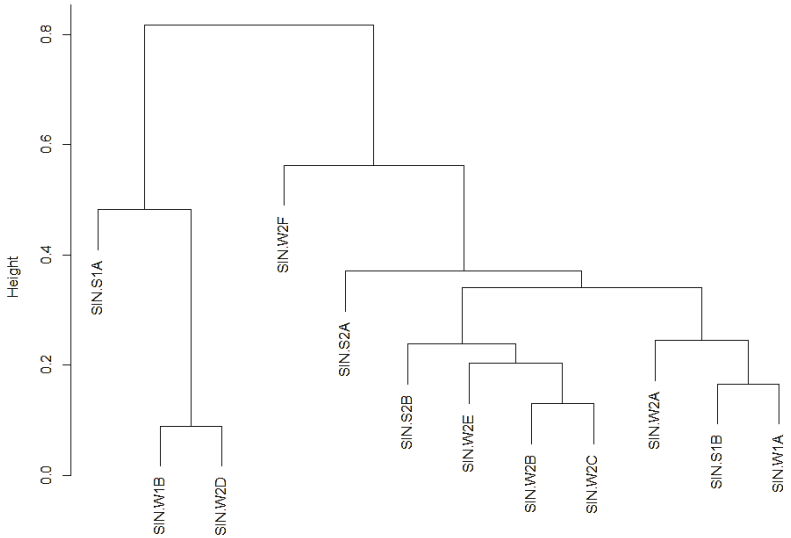


Figure 5.8.17. Cluster dendrogram of similarity across ICE text categories in SinE

It is clear from Figure 5.8.17, which displays the hierarchical cluster dendrogram (c = 0.82) for SinE, that in this variety the spoken text types are dispersed across the graphical representation and show no sign of

direct alignment as can be seen in the majority of the other varieties under investigation. The only significant cluster according to a post-hoc test with R is the rightmost one (s1b/w1a).

The latter point is confirmed by the non-hierarchical view in Figure 5.8.18. Not only is private dialogues (s1a) the text category with the lowest frequency of PrPfs (see above), it also emerges as the most noticeable outlier in the NeighborNet based on the comparison of all factors analyzed individually in the preceding sections. This once more emphasizes the particular position of this textual category, which is likely to show the strongest influence from the substrate via Singlish. The latter hypothesis seems to be borne out by the data; educated speakers of SinE are indeed also versatile in their usage according to the formality of the situation, as argued by Bao (2005: 264–265). The exact nature of this versatility will be further investigated below (Section 6.4).

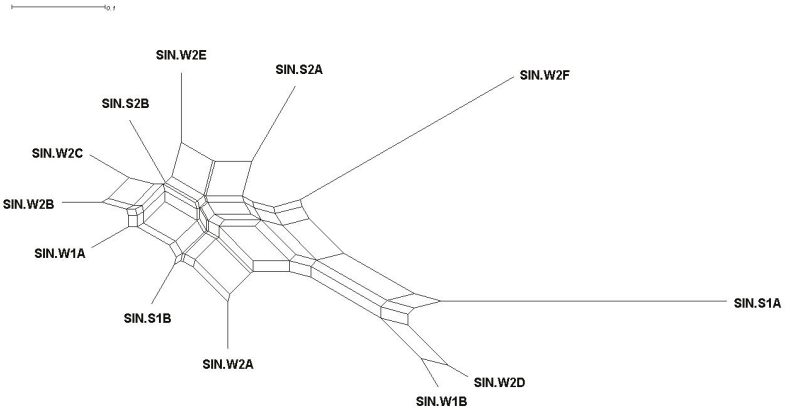


Figure 5.8.18. NeighborNet of similarity across ICE text categories in SinE

In addition, letters (w1b) and instructional (w2d) and creative (w2f) writing are distant from the remaining text types, which cluster toward the left of Figure 5.8.18. However, in respect of alignments of spoken and written texts, this cluster represents rather a "mixed bag", meaning that structural homogeneity of the spoken texts in the sense of Biber & Conrad (2009: 261) cannot be established based solely on the ICE-SIN data.

5.9 Hong Kong English (phase 3)

Another urban Southeast Asian variety of English, whose past has until recently been shaped by British colonial rule (for details see Görlach 2002: 104–107 or Setter et al. 2010: 103–112) and where structurally very different languages have been in contact, is HKE. Its socio-cultural history is comparable to SinE insofar that the foundation of Hong Kong is directly related to colonial activity. To be precise, British rule began after the first Opium War at the beginning of the 1840s with a strengthened position after the additional incorporation of the Kowloon peninsula in 1860 and again in 1898, when the "New Territories" were leased for 99 years. After a brief period of Japanese occupation during the Second World War, Hong Kong developed into one of the world's most globalized, cosmopolitan, economically active and therefore wealthiest cities, steadily growing through Chinese immigration. In accordance with the 1898 treaty, British rule ended after decade-long negotiations in 1997 (known as "the Handover") and Hong Kong has been one of the Special Administrative Regions of China since, a transitional status that will continue at least until 2047 and involves special legal, political and monetary regulations (Schneider 2007: 133–136; see further Evans 2009: 284–287 or Setter et al. 2010: 2–3 for demographic and sociolinguistic data).

Traditionally, English in Hong Kong was viewed as a language of oppression and elitism, as it was used by the colonial rulers and only accessible to a small part of the Chinese population strand, mainly through missionary schools. It has gained more and more currency especially after 1960, being associated with the strong economic growth and the emergence of what is known as the "new middle class" (Bolton 2003: 62; cited in Schneider 2007: 136) and related developments such as extended periods of study in L1 countries (Evans 2009: 281). At the same time, these developments interacted with changes in educational policy, which endorsed bilingualism for all (Evans 2009: 281), and led to a change of attitude toward English. The policy of bilingualism has, in practice, not been revised substantially by the Chinese government after 1997 (Bolton 2005: 9–10) and English is now considered a "value-added" (Li 1999; cited in Schneider 2007: 137) asset of the region. However, the

current linguistic situation is characterized by domain-specific usage and Chinese (mainly Cantonese and Putonghua) substrate influence in some areas: English is predominantly used in formal fields such as education and business but increasingly also in personal electronic communication and among workers. However, only in the past two decades (Bolton 2005: 3; cf. Köppl 1983b: 127) has HKE come to be accepted as a nativized variety with distinct characteristics on all structural levels (but see Ansaldo 2010: 513) and a simultaneous positive evaluation of HKE as well as of code-switching (known as "mix"; Schneider 2007: 138), also in educated circles.[264] Although this is not stated explicitly in the literature, we can safely assume that different lects are observable in actual usage also in Hong Kong, and this has led to continuing discussion on the status of HKE as an (emergent) L2 or rather learner variety (Görlach 2002: 109–110; Setter et al. 2010: 112–116). Some observers also claim increasing Americanization caused by cultural influences and a growing consumer culture oriented toward US models, although Bolton (2005: 4) suggests that the evidence for linguistic effects of these developments presently is not compelling.

Features of HKE grammar are comparatively unobtrusive (Ansaldo 2010: 499). Those relevant for the present study are comparable to the findings for a number of other L2 varieties (such as JamE or SinE). In particular, it has been noted that speakers of HKE occasionally rely on a TR system that merely uses past and present or only the latter, and that temporal adverbials are used instead to mark temporal and aspectual contrasts in colloquial settings (Platt 1982: 410; Kachru & Nelson 2006: 42; Setter et al. 2010: 49; see also Section 2.1), as is illustrated in (1).

(1) ask any question that you come across in the learning process on any topics or issue that we *already* cover during the last lecture okay (ICE-HK s1b-014)

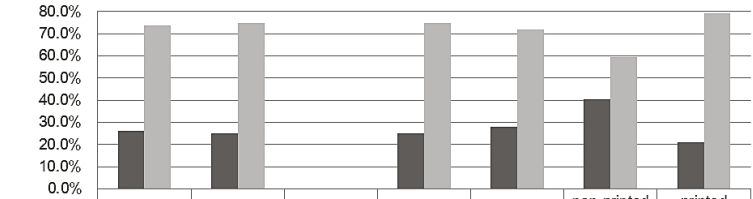
[264] Evans (2009: 297–298) suggests that "nativization" of HKE has started from the beginning of colonial rule already. In contrast, some authors argue that exonormative orientation toward BrE is still traceable (Schneider 2007: 137).

Also, invariant forms instead of past participles, as in (2) or (3), occur in the corpus data (see further Section 6.4).

(2) it has been shipping for almost twenty four months as of November ninety one and by July ninety one we have already *ship* over one million unit worldwide (ICE-HK s2a-022)
(3) I have just *listen* to them once. (ICE-HK w1b-009)

These findings suggest that HKE is a potential candidate for the lower end of the PrPf-friendliness spectrum, especially in informal speech and also in text types such as student writing, where speakers show more learner variety-like usage. However, the persistent exonormative orientation toward BrE mentioned above is a potential source overriding nativized patterns, in particular in the more formal domains. Therefore, a close look at text type differences is necessary here and, moreover, it should be considered to what extent speakers of HKE rely on temporal adverbials as additional perfect markers, potentially also with a higher share of specified PrPfs in involved text types.

5.9.1 Temporal adverbials

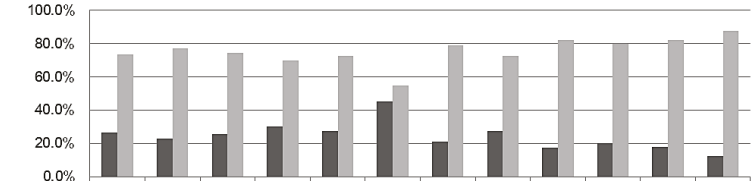


	SPOKEN (n=355)	WRITTEN (n=195)		dialogues (n=223)	monologues (n=132)	non-printed writing (n=42)	printed writing (n=153)
with TA	26.2%	25.1%		25.1%	28.0%	40.5%	20.9%
without TA	73.8%	74.9%		74.9%	72.0%	59.5%	79.1%

Figure 5.9.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-HK (register and macro-genre differences)

Figure 5.9.1 displays the relative frequencies of PrPf tokens that are specified by a temporal adverbial. In the ICE-HK data, overall more than one quarter of all instances is specified and differences between speech

and writing are not significant. Some variance is observable among the macro-genres, particularly in writing, in which printed writing, with approximately one in five, and non-printed writing, with approximately two in five PrPfs that are temporally specified, constitute the poles. Consequently, the two genres are significantly different from each other ($\chi^2 = 6.7022$, df = 1, $p < 0.01$, $\varphi_c = 0.185$) and the same applies when non-printed writing is compared to the overall ($\chi^2 = 4.2677$, df = 1, $p < 0.05$, $\varphi_c = 0.085$) and even the aggregated written ($\chi^2 = 4.0513$, df = 1, $p < 0.05$, $\varphi_c = 0.131$) breakdown. This outlier apart, none of the genres reveals any significant differences, so we can establish a mixed result with regard to the uniformity of the proportion of temporal specification in ICE-HK.



	HK S1A (n=144)	HK S1B (n=79)	HK S2A (n=59)	HK S2B (n=73)	HK W1A (n=11)	HK W1B (n=31)	HK W2A (n=33)	HK W2B (n=40)	HK W2C (n=34)	HK W2D (n=10)	HK W2E (n=28)	HK W2F (n=8)
with TA	26.4%	22.8%	25.4%	30.1%	27.3%	45.2%	21.2%	27.5%	17.6%	20.0%	17.9%	12.5%
without TA	73.6%	77.2%	74.6%	69.9%	72.7%	54.8%	78.8%	72.5%	82.4%	80.0%	82.1%	87.5%

Figure 5.9.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-HK (detailed text categories)

It was hypothesized above that the involved text types yield a greater proportion of specified tokens and indeed this seems to be borne out by the data, as the detailed analysis in Figure 5.9.2 shows. The shares of specified tokens in these text types exceed the values for all remaining (written) types apart from popular writing (w2b). In letters (w1b), nearly half of all tokens are temporally specified, while the value for creative writing (w2f; 12.5%) is the lowest in HKE overall.

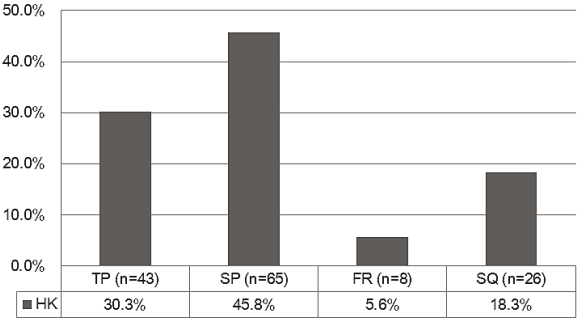


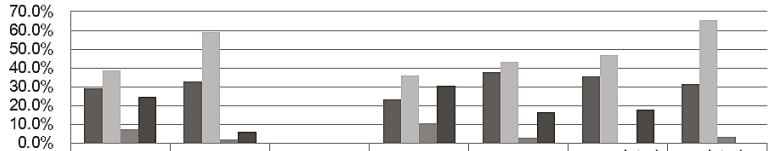
Figure 5.9.3. Types of PrPf adverbial specification: ICE-HK

As can be seen in Figure 5.9.3, which displays the breakdown of the different classes of temporal adverbial specification, those of span and duration (SP), as in (4), are once again the most frequent class, with more than two fifths of the relevant tokens, followed by adverbials of time-position (TP), as in (5), which cover nearly one third of the specified data. Adverbials of sequence (SQ), exemplified in (6), occur in less than one fifth of all cases, and frequency adverbials (FR), as in (7), are rare.

(4) You have been back *for half year* (ICE-HK s1a-096)
(5) *Only recently*, the humpback dolphins in this area have started to receive some attention from scientists. (ICE-HK w2b-027)
(6) We've been through that kind of school *before* (ICE-HK s1b-077)
(7) [...] I have met the man *on I would say at least two occasions* (ICE-HK s1b-031)

The breakdown of the adverbial classes in the two registers can be seen in Figure 5.9.4 below. It is apparent that, again, SP adverbials represent the most frequent category in both registers. However, SQ adverbials are comparatively prominent in speech (particularly in dialogues) and nearly as frequent as TP adverbials, while in writing FR and SQ adverbials are very rare. Thus, it is not surprising that the distributions of speech and writing are significantly different from each other (χ^2 = 10.861, df = 3,

$p < 0.05$, $\varphi_c = 0.277$). As concerns the macro-genres, the data are largely uniform. Printed writing can be identified as the only outlier whose breakdown is significantly different from the overall distribution ($\chi^2 = 8.267$, $df = 3$, $p < 0.05$, $\varphi_c = 0.218$).



	SPOKEN (n=93)	WRITTEN (n=49)		dialogues (n=56)	monologues (n=37)	non-printed writing (n=17)	printed writing (n=32)
TP	29.0%	32.7%		23.2%	37.8%	35.3%	31.3%
SP	38.7%	59.2%		35.7%	43.2%	47.1%	65.6%
FR	7.5%	2.0%		10.7%	2.7%	0.0%	3.1%
SQ	24.7%	6.1%		30.4%	16.2%	17.6%	0.0%

Figure 5.9.4. Types of PrPf adverbial specification: ICE-HK (register and macro-genre differences)

On the basis of the HKE data, we can form a hierarchy for the types of temporal specification that co-occur with PrPf tokens along the lines SP > TP > SQ > FR. Comparable to a number of other L2 varieties (e.g. PhiE and SinE), where previous literature has hinted at the use of SQ adverbials as markers of tense and aspect in colloquial speech, in ICE-HK, too, we find a comparatively high proportion of this adverbial type in dialogues. Further, the present analysis shows that the involved text categories are likely to yield a higher share of temporally specified PrPfs. This indicates that speakers of HKE rely on these arguably as additional TR markers, potentially through influence from the substrate and via the basilect.

5.9.2 Aktionsart

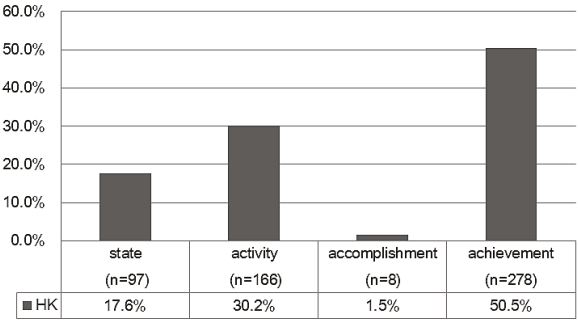


Figure 5.9.5. Aktionsart categories of main verbs in PrPf constructions: ICE-HK

In the ICE-HK sample, Aktionsart categories of PrPfs are distributed as shown in Figure 5.9.5. Half of the tokens are covered by achievement verbs, as exemplified in (8), slightly below one third by activity verbs, as exemplified in (9), while stative verbs, as in (10), cover nearly one sixth. Accomplishment verbs, as exemplified in (11), occur only sporadically.

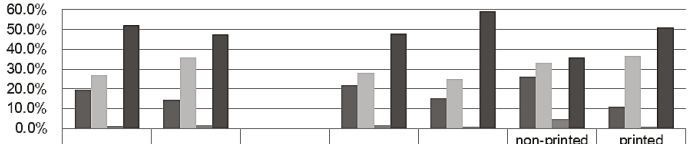
(8) And indeed for the Foundation they have not *found* uh uh pro problem by not be uh being one of the launching organizations (ICE-HK s1b-021)

(9) Containerization since the sixties has *gone* through stages of development (ICE-HK s2b-047)

(10) [...] Gibson mentions "Kowloon Walled City," which has "*continued* to haunt him" (n.pag.) ever since the latter told him about it. (ICE-HK w2a-010)

(11) They are friendly and I have *learned many things*. (ICE-HK w1b-015)

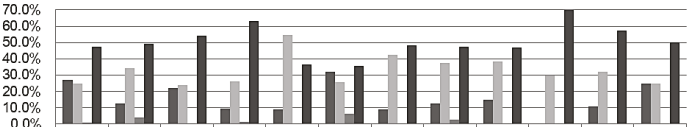
The proportions of the Aktionsart categories of PrPf tokens for the two modes of discourse are relatively uniform in ICE-HK (see Figure 5.9.6), a finding that is confirmed by a Pearson Chi-square test that returns a non-significant result. Achievements are the most common category in all macro-genres, while states are comparatively frequent in dialogues and non-printed writing, as are activities in the two written genres.



	SPOKEN (n=355)	WRITTEN (n=195)		dialogues (n=223)	monologues (n=132)	non-printed writing (n=42)	printed writing (n=153)
state	19.4%	14.4%		22.0%	15.2%	26.2%	11.1%
activity	27.0%	35.9%		28.3%	25.0%	33.3%	36.6%
accomplishment	1.4%	1.5%		1.8%	0.8%	4.8%	0.7%
achievement	52.1%	47.7%		48.0%	59.1%	35.7%	51.0%

Figure 5.9.6. Aktionsart categories of main verbs in PrPf constructions: ICE-HK (register and macro-genre differences)

All other tests for differences (macro-genres vs. overall/aggregated register proportions and register-internally) also yield non-significant results, apart from when the two written genres are compared ($\chi^2 = 10.5016$, $df = 3$, $p < 0.05$, $\varphi_c = 0.233$). Therefore, we can conclude that only minor effects of register or genre are present in the breakdown of Aktionsart categories of PrPf verbs in ICE-HK.



	HK S1A (n=144)	HK S1B (n=79)	HK S2A (n=59)	HK S2B (n=73)	HK W1A (n=11)	HK W1B (n=31)	HK W2A (n=33)	HK W2B (n=40)	HK W2C (n=34)	HK W2D (n=10)	HK W2E (n=28)	HK W2F (n=8)
state	27.1%	12.7%	22.0%	9.6%	9.1%	32.3%	9.1%	12.5%	14.7%	0.0%	10.7%	25.0%
activity	25.0%	34.2%	23.7%	26.0%	54.5%	25.8%	42.4%	37.5%	38.2%	30.0%	32.1%	25.0%
accomplishment	0.7%	3.8%	0.0%	1.4%	0.0%	6.5%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%
achievement	47.2%	49.4%	54.2%	63.0%	36.4%	35.5%	48.5%	47.5%	47.1%	70.0%	57.1%	50.0%

Figure 5.9.7. Aktionsart categories of main verbs in PrPf constructions: ICE-HK (detailed text categories)

At least two points worth noting are indicated by Figure 5.9.7, which shows the distributions of Aktionsart categories for the individual ICE text types. First, student writing (w1a) constitutes an outlier as it is the only type where activities are not the most frequent category, although it has to be acknowledged that the empirical basis for this finding is small (11 tokens only). A similar remark is in order for instructional writing

(w2d; 10 tokens), which is the only type without statives. Second, the spoken types are comparatively consistent (σ = 0.04), at least as far as the ranking of the categories (achievements > activities > states > accomplishments) is concerned (apart from the salience of state verbs in private dialogues (s1a)), while more variance (σ = 0.07) is observable among the written text categories.[265] This is in line with the results for a number of other L2 varieties (e.g. JamE).

5.9.3 Sentence type

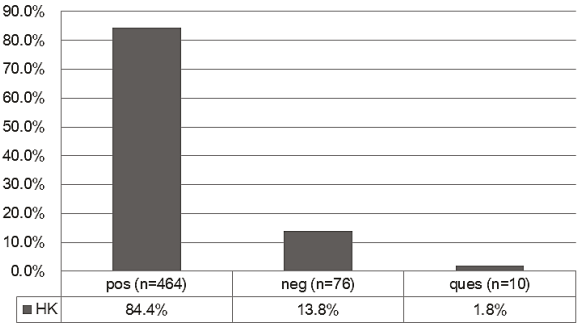
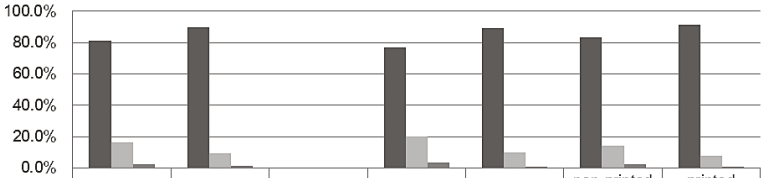


Figure 5.9.8 Distribution of sentence types: ICE-HK

Figure 5.9.8 shows that the vast majority of all tokens can be categorized as positive statements in the ICE-HK sample. Negatives, as illustrated in (12), occur in approximately every seventh token, and questions, as in (13), are rare.

(12) The English teaching establishment does not see this problem the way I do, probably because they haven't read about it in the huge, mostly foreign, literature on English-language teaching. (ICE-HK w2b-006)
(13) You you you've been to anywhere else (ICE-HK s1a-038)

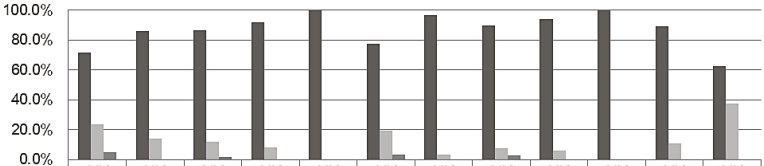
[265] This variance could to a certain extent be ascribed to the point that in the written categories token numbers are lower on average and therefore, individual occurrences exert a stronger influence on relative proportions.



	SPOKEN (n=355)	WRITTEN (n=195)		dialogues (n=223)	monologues (n=132)	non-printed writing (n=42)	printed writing (n=153)
pos	81.4%	89.7%		76.7%	89.4%	83.3%	91.5%
neg	16.3%	9.2%		20.2%	9.8%	14.3%	7.8%
ques	2.3%	1.0%		3.1%	0.8%	2.4%	0.7%

Figure 5.9.9. Distribution of sentence types: ICE-HK (register and macro-genre differences)

As can be seen from Figure 5.9.9, PrPfs with negatives and questions are approximately twice as frequent in speech compared to writing, while positives are clearly the majority variant in both modes of discourse. This makes the two registers significantly different from each other in that respect ($\chi^2 = 6.6812$, df = 2, $p < 0.05$, $\varphi_c = 0.11$). In the macro-genre analysis it emerges that both negatives and questions are most frequent in dialogues, which makes this genre significantly different from its spoken counterpart, monologues ($\chi^2 = 9.1493$, df = 2, $p < 0.05$, $\varphi_c = 0.161$), and when the overall breakdown is taken as a baseline ($\chi^2 = 6.5002$, df = 2, $p < 0.05$, $\varphi_c = 0.092$), albeit with a weak effect. In addition, printed writing approximates the 5% level of statistical significance when compared to the overall breakdown ($\chi^2 = 5.8302$, df = 2, $p = 0.054$, $\varphi_c = 0.091$).



	HK S1A (n=144)	HK S1B (n=79)	HK S2A (n=59)	HK S2B (n=73)	HK W1A (n=11)	HK W1B (n=31)	HK W2A (n=33)	HK W2B (n=40)	HK W2C (n=34)	HK W2D (n=10)	HK W2E (n=28)	HK W2F (n=8)
pos	71.5%	86.1%	86.4%	91.8%	100.0%	77.4%	97.0%	90.0%	94.1%	100.0%	89.3%	62.5%
neg	23.6%	13.9%	11.9%	8.2%	0.0%	19.4%	3.0%	7.5%	5.9%	0.0%	10.7%	37.5%
ques	4.9%	0.0%	1.7%	0.0%	0.0%	3.2%	0.0%	2.5%	0.0%	0.0%	0.0%	0.0%

Figure 5.9.10. Distribution of sentence types: ICE-HK (detailed text categories)

It emerges from Figure 5.9.10 that text categories further influence distributions of sentence types. The rare category of questions appears only in four text types (private dialogues (s1a); scripted monologues (s2a); letters (w1b); popular writing (w2b)), which are partly the same as in SinE. Note once more, however, that relative values here may be misleading when not interpreted cautiously, as they translate to one single token in each text type apart from private dialogues (s1a). Negatives do not occur at all in student (w1a) and persuasive (w2e) writing, while they are prominent in private dialogues (s1a) and creative writing (w2f), where direct speech or thought is imitated, as in (14) or (15):

(14) I haven't defrosted the chicken legs! (ICE-HK w2f-007)
(15) I haven't missed them, that's for sure. (ICE-HK w2f-012)

All in all, the data from the HKE sample indicate that text type may determine the distributions of sentence types in PrPf contexts. On a more general level of analysis, some effects of register and genre are also evident.

5.9.4 Semantics

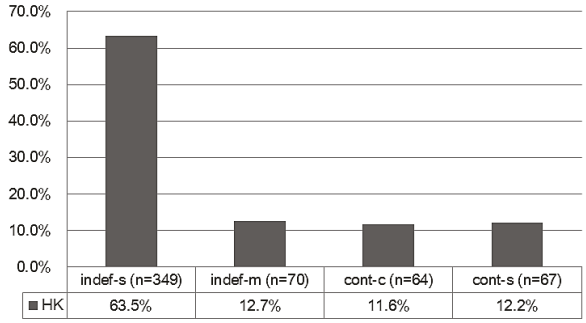


Figure 5.9.11. Distribution of semantic categories: ICE-HK

Figure 5.9.11 displays the overall breakdown of semantic categories of PrPfs in the HKE data and shows that indefinite single acts/events (indef-s), as in example (16), are highly frequent, with more than three out of five tokens. The share of the remaining categories is approximately

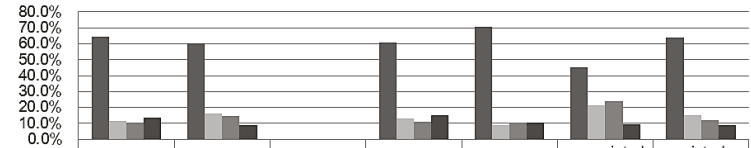
even, all covering around one eighth of the data. Strictly speaking, indefinite multiple acts (indef-m), as in (17), are slightly more common than continuatives acts/events (cont-c), as in (18), and continuative states (cont-s), as in example (19).

(16) THE right-wing Hongkong and Kowloon Trades Union Council has split over a call from its president to abandon its anti-communist stance. (ICE-HK w2c-009)

(17) Studies in Chinese infants have shown that most newborns with serum bilirubin concentrations above 221 mmol/L can be identified by using the jaundice meter, which gives a reading of > 20. (ICE-HK w2s-023)

(18) Maybe the concept is hard to understand for some whose countries have not experienced loss in natural treasures to foreign aggression in the past few centuries (ICE-HK s2b-040)

(19) The local kids, above, welcome visitors to what has become, with good reason, one of their country's premier tourist destinations (ICE-HK w2d-016)

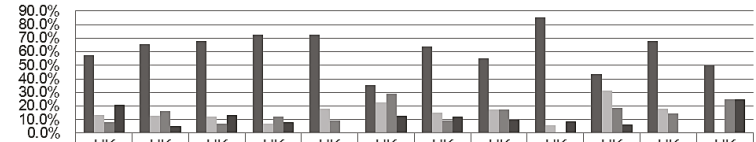


	SPOKEN (n=355)	WRITTEN (n=195)		dialogues (n=223)	monologues (n=132)	non-printed writing (n=42)	printed writing (n=153)
indef-s	64.2%	60.2%		60.5%	70.5%	45.2%	64.2%
indef-m	11.5%	16.4%		13.0%	9.1%	21.4%	15.1%
cont-c	10.7%	14.4%		11.2%	9.8%	23.8%	11.9%
cont-s	13.5%	9.0%		15.2%	10.6%	9.5%	8.8%

Figure 5.9.12. Distribution of semantic categories: ICE-HK (register and macro-genre differences)

A first glance at the relative values for the different semantic readings in the two registers, as presented in Figure 5.9.12, suggests no manifest dissimilarity and indeed, differences between the two are not statistically significant. The same result is found for the vast majority of other rele-

vant comparisons of macro-genres with the overall and register distributions as well as when compared register-internally, which suggests homogeneity of the data for this variable. One exception to this general pattern of ICE-HK is the result for non-printed writing. In this genre, indef-m and cont-c readings exceed one fifth of the PrPf tokens each, which renders its distribution significantly different from the overall breakdown ($\chi^2 = 8.1824$, df = 3, $p < 0.05$, $\varphi_C = 0.117$).



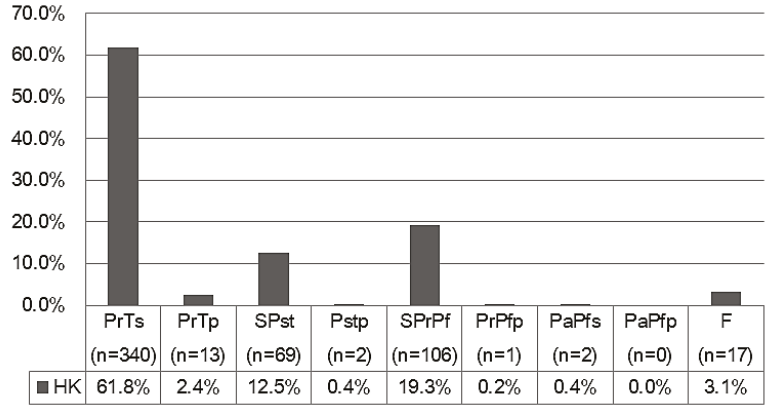
	HK S1A (n=144)	HK S1B (n=79)	HK S2A (n=59)	HK S2B (n=73)	HK W1A (n=11)	HK W1B (n=31)	HK W2A (n=33)	HK W2B (n=40)	HK W2C (n=34)	HK W2D (n=10)	HK W2E (n=28)	HK W2F (n=8)
indef-s	57.6%	65.8%	67.8%	72.6%	72.7%	35.5%	63.6%	55.0%	85.3%	43.8%	67.9%	50.0%
indef-m	13.2%	12.7%	11.9%	6.8%	18.2%	22.6%	15.2%	17.5%	5.9%	31.3%	17.9%	0.0%
cont-c	8.3%	16.5%	6.8%	12.3%	9.1%	29.0%	9.1%	17.5%	0.0%	18.8%	14.3%	25.0%
cont-s	20.8%	5.1%	13.6%	8.2%	0.0%	12.9%	12.1%	10.0%	8.8%	6.3%	0.0%	25.0%

Figure 5.9.13. Distribution of semantic categories: ICE-HK (detailed text categories)

As could be expected, more dispersion can be observed when the individual ICE text categories are taken into account (see Figure 5.9.13). Indef-s is the most frequent semantic reading in all types, but ranges of individual semantic categories vary considerably, particularly among the written texts. The overall variance in the latter is more than twice as high ($\sigma = 0.10$) as in the spoken categories ($\sigma = 0.04$). Specifically, both the lowest and the highest individual proportion for all four semantic categories can be allocated to a written text type and the text categories with empty cells are also all from writing. It could be supposed that this is merely due to lower token counts. In fact, this is not the case, as can be deduced from the values for reportage (w2c), a text category with 34 tokens but a conspicuous pattern, with the highest individual value for indef-s and an empty cell for cont-c.[266]

[266] On a related note, it emerges that the variance amongst the written categories still amounts to $\sigma = 0.09$ even if the values for reportage (w2c) or, alternatively, for letters (w1b; 31 tokens), another category that yields a comparatively idiosyncratic distributional pattern, are disregarded.

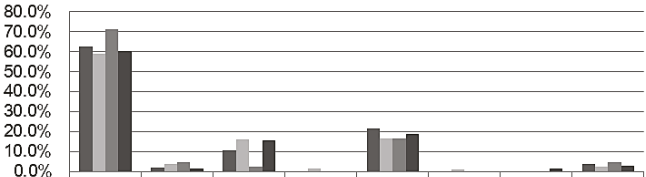
5.9.5 Preceding time reference



	PrTs (n=340)	PrTp (n=13)	SPst (n=69)	Pstp (n=2)	SPrPf (n=106)	PrPfp (n=1)	PaPfs (n=2)	PaPfp (n=0)	F (n=17)
HK	61.8%	2.4%	12.5%	0.4%	19.3%	0.2%	0.4%	0.0%	3.1%

Figure 5.9.14. TR forms preceding the PrPf: ICE-HK

The HKE sample also has only a limited set of preceding TR forms. It transpires from Figure 5.9.14 that PrTs, with more than three fifths of all tokens, is the most frequent option, while SPrPf occurs before nearly every fifth and SPst before every eighth PrPf form. All other TR forms apart from PrTp and F are very rare.



	PrTs (n=340)	PrTp (n=13)	SPst (n=69)	Pstp (n=2)	SPrPf (n=106)	PrPfp (n=1)	PaPfs (n=2)	F (n=17)
dialogues	62.8%	1.8%	10.3%	0.0%	21.5%	0.0%	0.0%	3.6%
monologues	59.1%	3.8%	15.9%	1.5%	16.7%	0.8%	0.0%	2.3%
non-printed writing	71.4%	4.8%	2.4%	0.0%	16.7%	0.0%	0.0%	4.8%
printed writing	60.1%	1.3%	15.7%	0.0%	19.0%	0.0%	1.3%	2.6%

Figure 5.9.15. TR forms preceding the PrPf: ICE-HK (macro-genre differences)

As regards the distribution of preceding TR forms in the macro-genres of the HKE data, Figure 5.9.15 reveals that the range of the values for the

highly frequent individual TR forms (PrTs, SPst and SPrPf) is fairly narrow. Further, differences between the spoken and written parts of the corpus are not significant (relative values not shown in Figure 5.9.15). Preceding SPst forms are comparatively uncommon in non-printed writing, but none of the tests for differences comparing the absolute macro-genre values to the overall and register breakdown and to the corresponding register-internal genre yields a significant result. This indicates homogeneous distributions as regards preceding TR forms in the present HKE sample.

5.9.6 Perfect-friendliness and text type effects

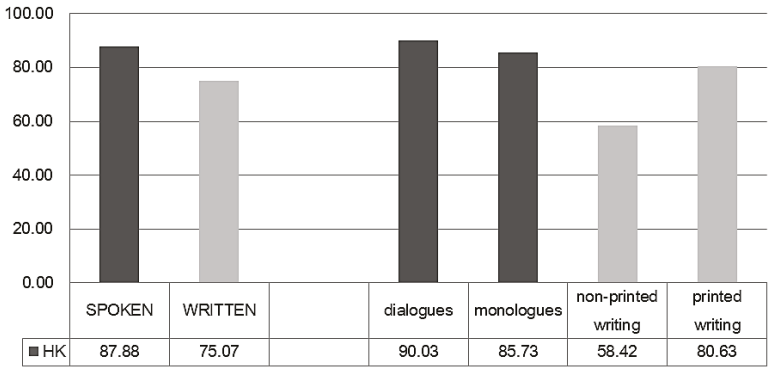


Figure 5.9.16. Frequency of PrPf occurrences in ICE-HK registers and macro-genres (normalized frequency per 10,000 words)

The PrPf-friendliness value (79 PrPfs/pttw) of the ICE-HK data is the second highest one overall, even exceeding the value for BrE. As in the majority of varieties, PrPfs occur more frequently in speech, while the difference between the two macro-genres that yield the most extreme values, dialogues (90) and non-printed writing (58), is immense (see Figure 5.9.16).

That data points are highly dispersed in HKE also becomes clear from the standard deviation across the text types (σ = 36.45), which represents the highest value of all varieties. In detail, PrPf-friendliness scores vary between 28 (student writing; w1a) and 168 (persuasive writ-

ing; w2e), which is the highest number for a single text type across the board. While the above hypotheses that HKE should be located at the lower end of the PrPf-friendliness scale overall and in informal speech are not confirmed by the data, learner effects are still very salient in student writing. Thus, the overall results for the HKE data can be viewed as illustrative of a variety-internal continuum of PrPf occurrence frequencies that is dependent on register and situation. Normalized PrPf frequencies correlate most strongly with the AusE values (r = 0.76), another variety that emerges as relatively PrPf-friendly.

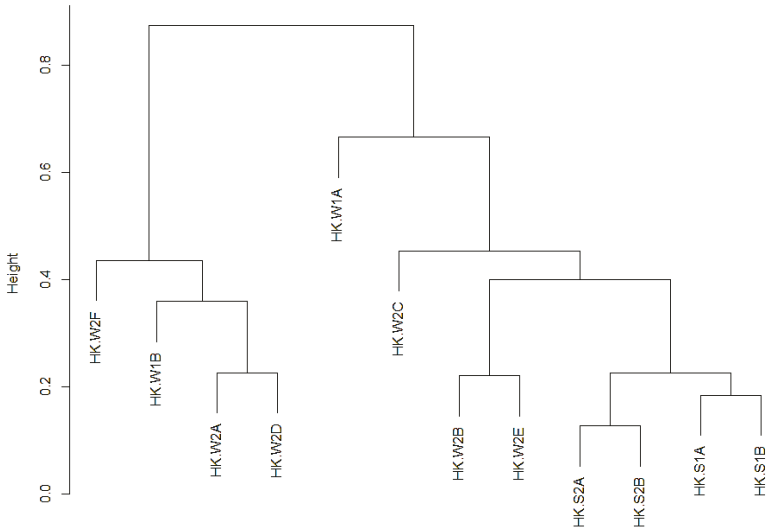


Figure 5.9.17. Cluster dendrogram of similarity across ICE text categories in HKE

The hierarchical cluster dendrogram (c = 0.73) displayed in Figure 5.9.17 shows that in this dataset the spoken text types are all located toward the right of the representation, with the monologue and dialogue pairs each forming neat clusters. Significance testing in R reveals that the monologue cluster, now including public dialogues (s1b), is significant.

The non-hierarchical display in Figure 5.9.18 confirms the grouping of the spoken text types. They are all located on neighboring nodes which share a number of splits. Letters (w1b) is the closest written type, while, once again, creative (w2f) and student (w1a) writing are distant

from the remaining textual categories; the latter arguably as it strongly shows learner effects, that is, the avoidance of the complex PrPf construction. Furthermore, we can conclude that the HKE data exemplify the close alignment of spoken texts as hypothesized by Biber & Conrad (2009: 261) in terms of structural homogeneity of the four categories as represented in ICE.

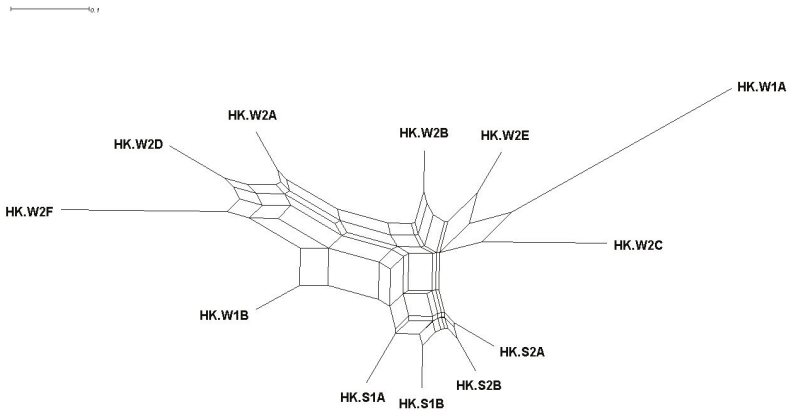


Figure 5.9.18. NeighborNet of similarity across ICE text categories in HKE

5.10 Indian English (phase 3)

English has been present in India for over 400 years and, in terms of sheer speaker numbers, IndE can be viewed as one of the largest varieties of English overall. The starting point of a more or less systematic relationship between India and Britain can be set in 1600, after which the East India Company established trade posts along the Indian coast. Gradually, political dominance of the area was also sought and the British Empire exerted control, first jointly with the East India Company (1784) and later on its own (1858). This went hand in hand with a stepwise transformation of India into an exploitation colony. Movements for liberation (e.g. the "Swadeshi Movement" or Ghandi's activities) increased in strength from the beginning of the 20th century onwards and

eventually resulted in independence in 1947, when the Indian subcontinent was partitioned along religious lines into India and Pakistan. A constitution was introduced in 1950 and India has since developed into one of the biggest and most economically thriving democracies worldwide, while many social issues and the political rivalry with Pakistan remain unsettled (Schneider 2007: 161–167).

The main linguistic input to early IndE were Cockney and the Northern English dialects of mostly uneducated speakers. Missionary schools are recognized as the first vehicle to spread the language, but more and more English-medium schools were established in the course of time, particularly in urban regions. Approximately from the beginning of the 19th century, both the indigenous elite and the colonial rulers acknowledged the benefits of a layer of bilingual Indians, albeit for different purposes (see Brutt-Griffler 2002: 39–60 for discussion), and a bilingual language policy was officially endorsed, later also including education of the middle class. Thus, it does not come as a surprise that English was still considered elitist and a sign of colonial oppression after independence and was to be replaced by Hindi as a national language. However, this never materialized for various reasons (see Dua 1996: 560–564 for details) and in 1967 English was recognized as a co-official language in what is known as the "Three Language Formula" (see Vaish 2008: 12–15). Currently, IndE is the native language of a small percentage of the Indian population only, while it mainly fulfills a utilitarian (typically, L2) function for the majority as ethnically neutral and economically helpful language. Therefore, it has been labeled a "non-native institutionalized variety" (Sharma 2001: 344), as, in terms of (earlier) language contact, its speakers have "not [been] integrated into the original TL (= target language; V.W.) speech community" (Thomason 2001: 75). However, what distinguishes it from a number of other Asian varieties of English is the fact that it cannot be categorized as an identity marker for its speakers. Along with a continuing strong presence of indigenous languages according to the Three Language Formula, this is mainly due to a persistent exonormative orientation toward BrE mod-

els,[267] even if nativized patterns and some minor indications of endonormative stabilization are clearly identifiable (Schneider 2007: 169–173).

In the area of grammar, and for PrPf usage in colloquial contexts in particular, some nativized preferences are salient. Above all, studies have found an overuse/extended functional range of the PrPf with some speakers of IndE. It can occur in environments where typically the SPst or the Past Perfect would be used (McArthur 2002: 322; Trudgill & Hannah 2008: 137; Schneider & Hundt 2012: 23; Werner 2013b: 230; see also sections 6.4 and 6.5). Others, however, have doubted that IndE patterns differently from BrE and AmE in this respect (Balasubramanian 2009: 92–93).[268] In addition, it has been found that, (i) similarly to other Asian varieties, perfect marking is sometimes absent and present forms and/or temporal adverbials (such as *already*) are used for marking tense and aspect relations instead (Sharma 2001: 367; Bhatt 2004: 1028; cf. Sharma 2012) and that (ii) occasionally, reduced forms (such as *I done it*) are in evidence (Davydova 2011: 181). Many of these features are nicely illustrated in the following examples (see further Section 6.4):

(1) I never *beat* a student never my whole life (ICE-IND s1a-085)
(2) Then they do not bother about the taste they have just *eat* it and *rush* like that (ICE-IND s1a-077)

Additionally, the different sequence of tense rules of the respective indigenous languages may exert an influence on the use of TR forms in IndE (Kachru & Smith 2008: 93–94).

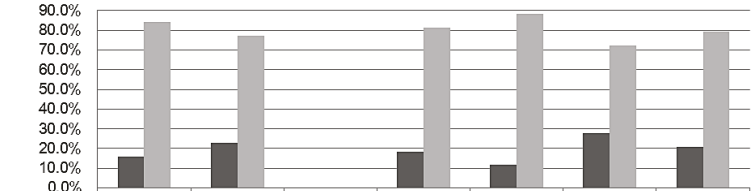
We can derive two main points from previous studies: First, due to the persistent exonormative orientation toward BrE, it is likely that the ICE-IND data yield PrPf-friendliness scores toward the high end of the

[267] Sailaja (2009: 39–43) remarks that AmE influence (mainly through popular culture) might be increasing in the future. Note, however, that Dua (1996: 578ff) has diagnosed a continuing strong presence of the British Council in India, which in turn suggests persistence of BrE norms in education.

[268] Quantitatively, this may be true. Qualitative differences in the cases where PrPfs are used in typical SPst contexts (e.g. with definite temporal adverbials) do exist, however (see Werner 2013b).

hierarchy, potentially even exceeding the values for BrE in some text types, as an extended functional range of the PrPf has been identified for IndE. However, some nativized features or patterns attributable to learner effects may be salient in colloquial texts, so careful scrutiny of the influence of register, genre and text type is indispensable. Second, like in HKE, we should expect a high proportion of temporal specification in involved text types and determine the exact nature of that temporal specification.

5.10.1 Temporal adverbials

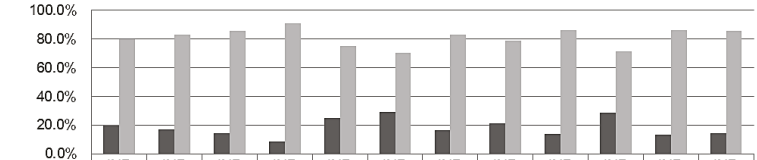


	SPOKEN (n=388)	WRITTEN (n=190)		dialogues (n=227)	monologues (n=161)	non-printed writing (n=54)	printed writing (n=136)
with TA	15.7%	22.7%		18.5%	11.8%	27.8%	20.9%
without TA	84.3%	77.3%		81.5%	88.2%	72.2%	79.1%

Figure 5.10.1. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-IND (register and macro-genre differences)

The values for PrPf tokens in ICE-IND that are temporally specified by an adverbial are shown in Figure 5.10.1. Overall, fewer than one in five PrPfs are specified and the value for speech is particularly low compared to all other varieties under investigation. Accordingly, the difference between the aggregated spoken and written parts of the corpus component is significant ($\chi^2 = 4.4313$, df = 1, $p < 0.05$, $\varphi_C = 0.086$). Within the former register, monologues is the macro-genre with the lowest specification value and also the only genre that borders the 5% level of statistical significance when compared to the overall distribution ($\chi^2 = 3.6556$, df = 1, $p = 0.056$, $\varphi_C = 0.07$). Note, however, that effect sizes for both established significant differences are small. With specified tokens exceeding one quarter, non-printed writing is the macro-genre with the highest proportion of specified PrPfs. However, it is not significantly

different from the overall distribution, when compared to the respective register breakdown or register-internally. The same applies to all remaining macro-genres, which suggests relative homogeneity of the ICE-IND data as regards temporal specification values.



	IND S1A (n=110)	IND S1B (n=117)	IND S2A (n=83)	IND S2B (n=78)	IND W1A (n=20)	IND W1B (n=34)	IND W2A (n=24)	IND W2B (n=13)	IND W2C (n=29)	IND W2D (n=14)	IND W2E (n=22)	IND W2F (n=14)
with TA	20.0%	17.1%	14.5%	9.0%	25.0%	29.4%	16.7%	21.2%	13.8%	28.6%	13.6%	14.3%
without TA	80.0%	82.9%	85.5%	91.0%	75.0%	70.6%	83.3%	78.8%	86.2%	71.4%	86.4%	85.7%

Figure 5.10.2. Proportion of PrPf occurrences with and without temporal specification by adverbial(s): ICE-IND (detailed text categories)

A more diverse picture emerges from Figure 5.10.2, which displays the proportions for the individual ICE text types. As could be inferred from the preceding macro-genre analysis, temporally specified PrPfs are comparatively frequent in the involved (i.e. non-printed) written text types. Furthermore, instructional writing (w2d) shows a specification value well above average. Remarkably, the text type with the lowest values, scripted monologues (s2b), can be found among the spoken categories. Private dialogues (s1a) is the only spoken type whose specification value borders the written average, and this is quite striking in light of the findings of previous studies on IndE (see above) and on other varieties (e.g. HKE), which suggested high(er) shares of specified tokens in the spoken categories. In the ICE-IND data, the usual correlation of spoken texts with a colloquial style (and the accordingly higher amounts of PrPf specification) seems to be suspended or at least weakened.

Figure 5.10.3 below displays the relative frequencies of the various types of temporal adverbials that occur in those PrPf tokens that are specified. Adverbials of span and duration (SP), as in example (3), are the majority variant, with more than two fifths of all instances. Adverbials of time-position (TP), exemplified in (4), cover one third and those of sequence (SQ), exemplified in (5), less than one fifth of the relevant data.

Frequency adverbials (FR), as in (6), again represent the least common class.

(3) Well taking of hostages has been with us *for sometimes* (ICE-IND s1b-024)
(4) *That day* also he has told us that in the middle of the copper wire we put a magnet (ICE-IND s1b-019)
(5) [...] you have informed us *earlier* that 'Mathematical Education' vol.7 no. 3-4 are still in press and you will send these issues [...]. (ICE-IND w1b-027)
(6) And one of the uh interesting thing that that happen in this Davis Cup ties I think that India has won the toss *in every match* (ICE-IND s2a-004)

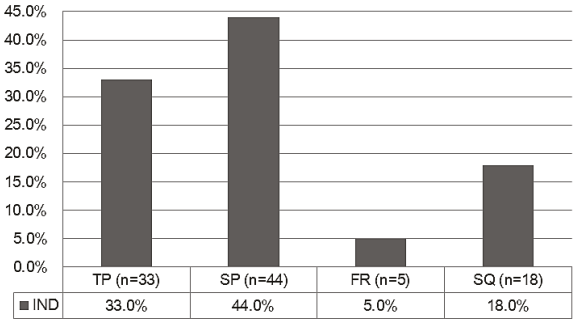
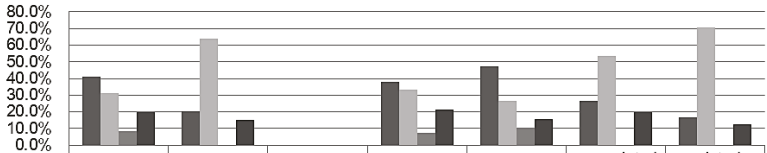


Figure 5.10.3. Types of PrPf adverbial specification: ICE-IND

The proportions of the different types of adverbial specification, shown in Figure 5.10.4, diverge quite considerably, with TP, FR and SQ adverbials being more frequent in speech than in writing. Additionally, SP is not the most common type in the former register. Not unexpectedly, a Pearson Chi-square test shows that differences between the registers are statistically significant ($\chi^2 = 12.333$, $df = 3$, $p < 0.01$, $\varphi_c = 0.351$). Although the relative values suggest otherwise at first glance, there are no outliers among the macro-genres (i.e. none of the tests for differences based on the absolute numbers yields a statistically significant result), which indicates homogeneity of the distributions. Note, however, that the results of the present analysis have to be interpreted with great care,

as the absolute counts are low. SQ adverbials are slightly more prominent in dialogues, while the rare category of FR adverbials is completely absent from the two written macro-genres.



	SPOKEN (n=61)	WRITTEN (n=39)		dialogues (n=42)	monologues (n=19)	non-printed writing (n=15)	printed writing (n=24)
TP	41.0%	20.5%		38.1%	47.4%	26.7%	16.7%
SP	31.1%	64.1%		33.3%	26.3%	53.3%	70.8%
FR	8.2%	0.0%		7.1%	10.5%	0.0%	0.0%
SQ	19.7%	15.4%		21.4%	15.8%	20.0%	12.5%

Figure 5.10.4. Types of PrPf adverbial specification: ICE-IND (register and macro-genre differences)

In the ICE-IND sample, the overall hierarchy for the types of temporal specification of PrPfs goes along the lines SP > TP > SQ > FR. Similarly to other Asian varieties of English (HKE, PhiE, SinE), we can observe a more pronounced usage of SQ adverbials in dialogues, which may be due to substrate influence (see above), although this tendency is not as clear as in the other varieties. A similar finding applies to the hypothesis that PrPf tokens in involved texts are more likely to be temporally specified. In the ICE-IND data, only the second hypothesis can be corroborated, while speech, apart from private dialogues (s1a), does not seem to rely on extensive temporal specification as could have been supposed from previous studies.[269]

5.10.2 Aktionsart

Figure 5.10.5 below displays the breakdown of the Aktionsart types of PrPf tokens in the ICE-IND data sample. It shows that achievement verbs, as illustrated in (7), are highly frequent, with nearly three out of

[269] It is clear that basilectal and lower mesolectal data would probably contain higher rates of specification (for a qualitative analysis of examples see Section 6.5 and Werner 2013b).

five tokens. The relative frequency of activity verbs, as in (8), amounts to close to one in four and of state verbs, as in (9), to approximately one out of seven. Accomplishments, as illustrated in (10), are rare once again.

(7) Our 'Third Generation Computer' (Made in Heaven!) has *arrived*! (ICE-IND w1b-005)
(8) When the assessee has *followed* a particular method of accounting and has been regularly following it, the Income-tax Department must have sufficient material for rejecting the method of accounting. (ICE-IND w2a-015)
(9) Since time immemorial human society has *felt* the need for curbing its growth and this has led to the development of various contraceptive methods. (ICE-IND w2d-009)
(10) Smile to yourself until you have *warmed your own heart* with the sunshine of the cheerful continence that go out radiant your smile (ICE-IND s1a-001)

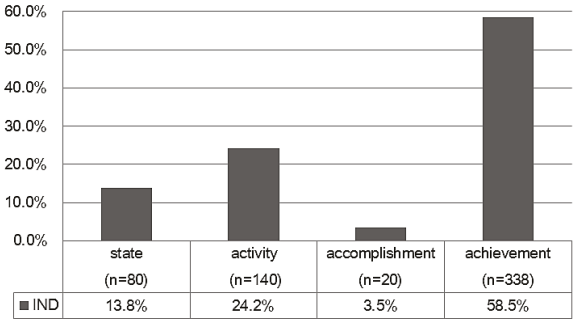
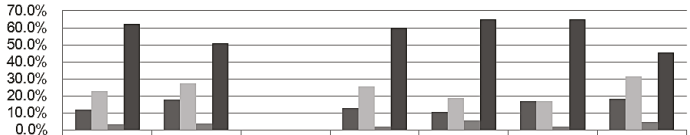


Figure 5.10.5. Aktionsart categories of main verbs in PrPf constructions: ICE-IND

Some dispersion is in evidence with regard to the relative shares of the Aktionsart categories in IndE, as shown in Figure 5.10.6. However, the distributions for speech and writing are not significantly different from each other. Likewise, the macro-genre analysis shows relative uniformity, while activity verbs are noticeably more common in printed writing. The latter therefore is the only genre where a Pearson Chi-square test borders the 5% level of significance when its distribution is compared to the overall breakdown ($\chi^2 = 7.4325$, df = 3, p = 0.059, $\varphi_c = 0.102$). In sum,

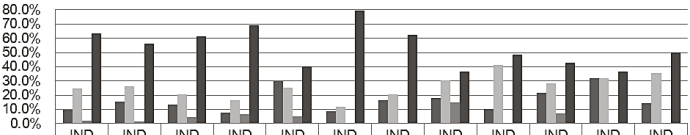
no strong effects of register or genre on the distributions of the Aktionsart classes of PrPf verbs are in evidence.



	SPOKEN (n=388)	WRITTEN (n=190)		dialogues (n=227)	monologues (n=161)	non-printed writing (n=54)	printed writing (n=136)
state	11.9%	17.9%		12.8%	10.6%	16.7%	18.4%
activity	22.7%	27.4%		25.6%	18.6%	16.7%	31.6%
accomplishment	3.4%	3.7%		1.8%	5.6%	1.9%	4.4%
achievement	62.1%	51.1%		59.9%	65.2%	64.8%	45.6%

Figure 5.10.6. Aktionsart categories of main verbs in PrPf constructions: ICE-IND (register and macro-genre differences)

A pattern that is not unexpected in light of the results for a number of other Asian varieties (e.g. HKE) can be seen in Figure 5.10.7. The distributions of the spoken text types are relatively consistent ($\sigma = 0.03$), while more variance occurs in the written text types ($\sigma = 0.07$). Yet, while frequencies vary, achievement verbs are the most frequent Aktionsart category across the board and the ranking of the categories (achievements > activities > states > accomplishments) is also similar in all texts.



	IND S1A (n=110)	IND S1B (n=117)	IND S2A (n=83)	IND S2B (n=78)	IND W1A (n=20)	IND W1B (n=34)	IND W2A (n=24)	IND W2B (n=13)	IND W2C (n=29)	IND W2D (n=14)	IND W2E (n=22)	IND W2F (n=14)
state	10.0%	15.4%	13.3%	7.7%	30.0%	8.8%	16.7%	18.2%	10.3%	21.4%	31.8%	14.3%
activity	24.5%	26.5%	20.5%	16.7%	25.0%	11.8%	20.8%	30.3%	41.4%	28.6%	31.8%	35.7%
accomplishment	1.8%	1.7%	4.8%	6.4%	5.0%	0.0%	0.0%	15.2%	0.0%	7.1%	0.0%	0.0%
achievement	63.6%	56.4%	61.4%	69.2%	40.0%	79.4%	62.5%	36.4%	48.3%	42.9%	36.4%	50.0%

Figure 5.10.7. Aktionsart categories of main verbs in PrPf constructions: ICE-IND (detailed text categories)

An exception is student writing (w1a), where state verbs are the second most frequent category. This finding relates to an observation from an earlier study that found state and activity verbs to be highly frequent

Aktionsart categories with the PrPf in learner English (Liszka 2003: 18) and thus suggests the presence of a learner effect, which seems plausible for this text category in particular.

5.10.3 Sentence type

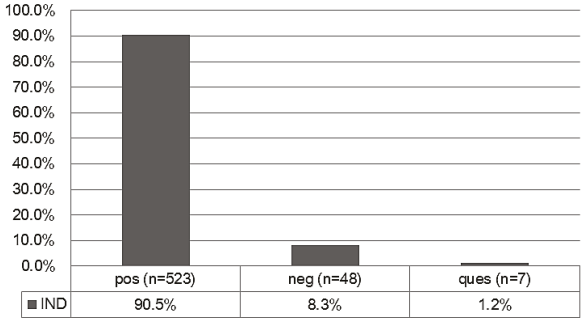
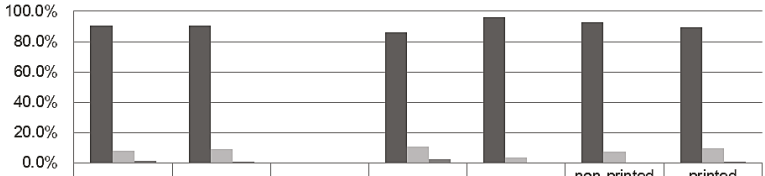


Figure 5.10.8. Distribution of sentence types: ICE-IND

More than nine out of ten PrPfs in the ICE-IND data occur in positive statements (see Figure 5.10.8). The remaining PrPfs surface in negative sentences and a small portion in questions, as examples (11) and (12) illustrate.

(11) The report is very clear that we have not mentioned any life member because we have uh stopped enrolling life members since nineteen seventy-four (ICE-IND s1b-080)
(12) So you have started dealing with your patients and all? (ICE-IND s1a-090)

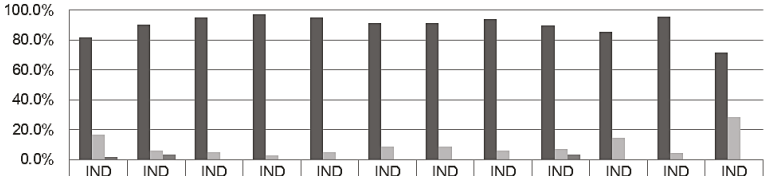
Figure 5.10.9 shows that the relative frequencies for the sentence types in which PrPfs occur do not diverge to a great extent between the two modes of discourse and, indeed, the differences are not significant. More variance is present in the macro-genres. The number of questions in dialogues is twice the overall figure, but they do not surface in monologues and non-printed writing.



	SPOKEN (n=388)	WRITTEN (n=190)		dialogues (n=227)	monologues (n=161)	non-printed writing (n=54)	printed writing (n=136)
pos	90.5%	90.5%		86.3%	96.3%	92.6%	89.7%
neg	8.0%	8.9%		11.0%	3.7%	7.4%	9.6%
ques	1.5%	0.5%		2.6%	0.0%	0.0%	0.7%

Figure 5.10.9. Distribution of sentence types: ICE-IND (register and macro-genre differences)

In monologues, PrPfs in negative sentences are also comparatively rare; PrPfs almost exclusively occur in positive sentences, making monologues a conspicuous genre. This impression is confirmed by Pearson Chi-square tests, which reveal that the breakdown for monologues is significantly different from both the overall ($\chi^2 = 5.9745$, df = 2, $p < 0.05$, $\varphi_C = 0.09$) and the written ($\chi^2 = 5.9745$, df = 2, $p = 0.050$, $\varphi_C = 0.104$) distribution as well as when compared register-internally with dialogues ($\chi^2 = 11.541$, df = 2, $p < 0.01$, $\varphi_C = 0.172$). Given, however, that effect sizes are partly weak and that no other significant differences are evident when register and macro-genre distributions are contrasted, it appears that only some minor effects of register and genre on the distribution of sentence types where PrPf tokens occur are present in ICE-IND.



	IND S1A (n=110)	IND S1B (n=117)	IND S2A (n=83)	IND S2B (n=78)	IND W1A (n=20)	IND W1B (n=34)	IND W2A (n=24)	IND W2B (n=13)	IND W2C (n=29)	IND W2D (n=14)	IND W2E (n=22)	IND W2F (n=14)
pos	81.8%	90.6%	95.2%	97.4%	95.0%	91.2%	91.7%	93.9%	89.7%	85.7%	95.5%	71.4%
neg	16.4%	6.0%	4.8%	2.6%	5.0%	8.8%	8.3%	6.1%	6.9%	14.3%	4.5%	28.6%
ques	1.8%	3.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.4%	0.0%	0.0%	0.0%

Figure 5.10.10. Distribution of sentence types: ICE-IND (detailed text categories)

Text types also exert some influence on the distributions, as Figure 5.10.10 reveals. Questions occur in three types only. As could be expected, these are the two dialogue types (s1a; s1b) as well as reportage (w2c). Note, however, that the relative value for the last text type translates into a single token, which is arguably used for some kind of rhetorical purpose:

(13) But, will this sudden burst of enthusiasm be optimally utilized by the Board, which at best, has remained non-functional? (ICE-IND w2c-019)

Negatives are particularly salient in private dialogues (s1a), instructional (w2d) and, most markedly, creative (w2f) writing, where again direct speech and thought is imitated.

5.10.4 Semantics

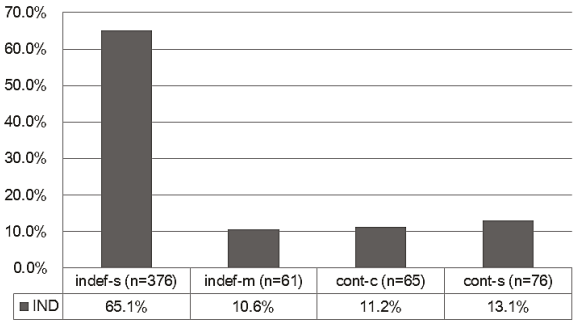
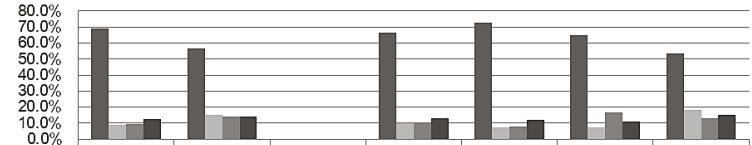


Figure 5.10.11. Distribution of semantic categories: ICE-IND

From the overall proportions of the four semantic categories in ICE-IND, presented in Figure 5.10.11, we can see that indefinite single act/event readings (indef-s), as in (14), with close to two thirds of all PrPfs are again the majority variant. Continuative state readings (cont-s), as in (15), are the second most frequent category, while the shares of continuative acts/events (cont-c), as in (16), and indefinite multiple acts (indef-m), as in (17), are virtually equivalent, with approximately one in nine tokens each.

(14) In Rajasthan the striking lawyers have decided to call off their state wide agitation from tomorrow (ICE-IND s2b-009)
(15) Important artistic movements in our century, from surrealism, expressionism and futurism, to abstract art, have felt the need to have a dialogue with cinema. (ICE-IND w2b-006)
(16) I have worked so hard to fulfil that dream and now you are telling me that you won't go for your specialisation! (ICE-IND w2f-015)
(17) In some sense uh even countries which don't have uh planning in the sense that we have used this word uh still have to plan their micro-economy (ICE-IND s1b-026)

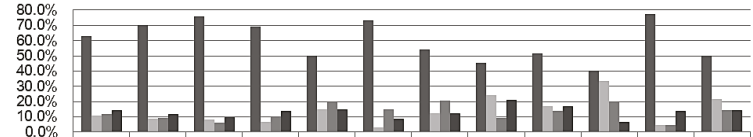


	SPOKEN (n=388)	WRITTEN (n=190)		dialogues (n=227)	monologues (n=161)	non-printed writing (n=54)	printed writing (n=136)
indef-s	69.1%	56.5%		66.5%	72.7%	64.8%	53.3%
indef-m	8.8%	15.2%		9.7%	7.5%	7.4%	18.2%
cont-c	9.5%	14.1%		10.6%	8.1%	16.7%	13.1%
cont-s	12.6%	14.1%		13.2%	11.8%	11.1%	15.3%

Figure 5.10.12. Distribution of semantic categories: ICE-IND (register and macro-genre differences)

Figure 5.10.12 displays the breakdown of the semantic categories for the registers and macro-genres in the IndE sample. It reveals that the highly frequent indef-s category is less prominent, while all other classes are consistently more common in writing and, on aggregate, these differences clearly reach the level of statistical significance ($\chi^2 = 10.614$, df = 3, $p < 0.05$, $\varphi_C = 0.135$). A closer look at the macro-genres indicates that the dispersion of the percentages for individual semantic categories among the genres is not very marked but that printed writing stands out in that its value for indef-m is at least twice as high as in the other genres. Accordingly, a Pearson Chi-square test based on the absolute numbers shows that its breakdown is significantly different from the overall breakdown ($\chi^2 = 9.9458$, df = 3, $p < 0.05$, $\varphi_C = 0.119$). Thus, with regard to

the homogeneity of the data, some register and genre effects on the distribution of the semantic categories can be observed in ICE-IND.



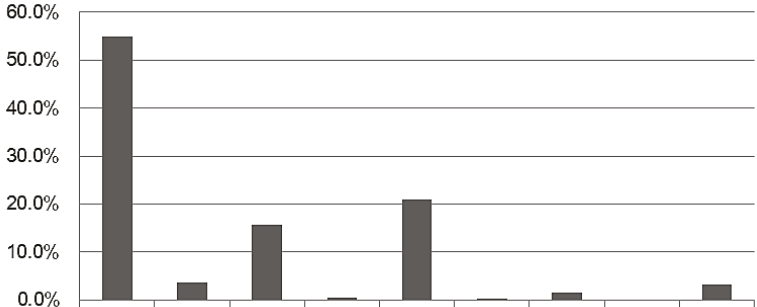
	IND S1A (n=110)	IND S1B (n=117)	IND S2A (n=83)	IND S2B (n=78)	IND W1A (n=20)	IND W1B (n=34)	IND W2A (n=24)	IND W2B (n=13)	IND W2C (n=29)	IND W2D (n=14)	IND W2E (n=22)	IND W2F (n=14)
indef-s	62.7%	70.1%	75.9%	69.2%	50.0%	73.5%	54.2%	45.5%	51.7%	40.0%	77.3%	50.0%
indef-m	10.9%	8.5%	8.4%	6.4%	15.0%	2.9%	12.5%	24.2%	17.2%	33.3%	4.5%	21.4%
cont-c	11.8%	9.4%	6.0%	10.3%	20.0%	14.7%	20.8%	9.1%	13.8%	20.0%	4.5%	14.3%
cont-s	14.5%	12.0%	9.6%	14.1%	15.0%	8.8%	12.5%	21.2%	17.2%	6.7%	13.6%	14.3%

Figure 5.10.13. Distribution of semantic categories: ICE-IND (detailed text categories)

The detailed view of the percentages of the semantic categories in the various ICE text types, as presented in Figure 5.10.13, shows that indef-s readings are consistently the most frequent variant. However, there is considerable variance among the percentages. All extreme values (e.g. the high proportion of indef-m and the small share of cont-s in instructional writing (w2d)) can be found within the written text types. That the distributions in the spoken text types are more uniform is further illustrated by the fact that the overall variance is nearly three times as high in writing ($\sigma = 0.08$) compared to the aggregated average value for the spoken texts ($\sigma = 0.03$).

5.10.5 Preceding time reference

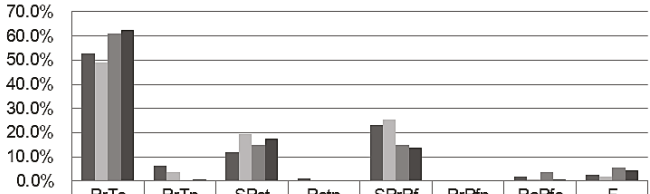
It is apparent from Figure 5.10.14 that in the IndE data only a restricted set of preceding TR forms occurs in noticeable quantities. Specifically, PrTs covers more than half of all tokens, while the share of SPrPf forms exceeds one fifth and SPst forms surface before fewer than every sixth token. The proportions of the remaining preceding TR forms, apart from PrTp and F, are negligible.



	PrTs (n=317)	PrTp (n=21)	SPst (n=90)	Pstp (n=2)	SPrPf (n=121)	PrPfp (n=1)	PaPfs (n=8)	PaPfp (n=0)	F (n=18)
IND	54.8%	3.6%	15.6%	0.3%	20.9%	0.2%	1.4%	0.0%	3.1%

Figure 5.10.14. TR forms preceding the PrPf: ICE-IND

Figure 5.10.15 reveals some variance in the percentages of the highly frequent TR forms, PrTs, SPst, and SPrPf. None of the distributions of the individual macro-genres is significantly different from the overall and the respective register breakdown or when compared register-internally. However, the aggregated register breakdowns are significantly different from each other ($\chi^2 = 20.587$, df = 7, $p < 0.01$, $\varphi_c = 0.189$; relative frequencies not shown in Figure 5.10.15). Thus, the overall uniformity of these distributions in the ICE-IND data is somewhat restricted by register effects.



	PrTs (n=317)	PrTp (n=21)	SPst (n=90)	Pstp (n=2)	SPrPf (n=121)	PrPfp (n=1)	PaPfs (n=8)	F (n=18)
dialogues	52.9%	6.2%	11.9%	0.9%	23.3%	0.4%	1.8%	2.6%
monologues	49.1%	3.7%	19.3%	0.0%	25.5%	0.0%	0.6%	1.9%
non-printed writing	61.1%	0.0%	14.8%	0.0%	14.8%	0.0%	3.7%	5.6%
printed writing	62.5%	0.7%	17.6%	0.0%	14.0%	0.0%	0.7%	4.4%

Figure 5.10.15. TR forms preceding the PrPf: ICE-IND (macro-genre differences)

5.10.6 Perfect-friendliness and text type effects

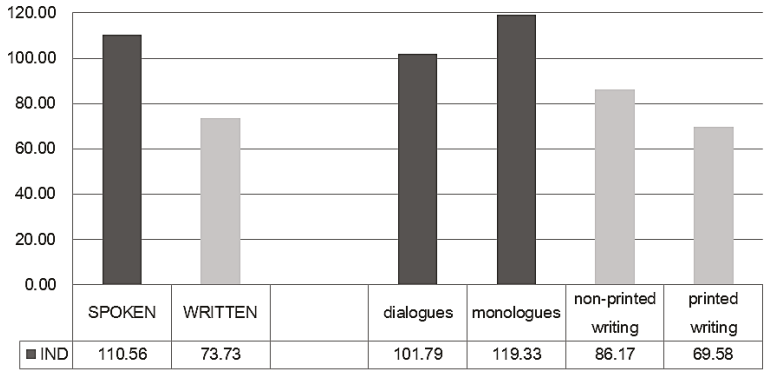


Figure 5.10.16. Frequency of PrPf occurrences in ICE-IND registers and macro-genres (normalized frequency per 10,000 words)

Among the varieties under investigation, IndE has the second highest overall PrPf-friendliness score (86 PrPfs/pttw), a finding which is in line with previous research that has found high usage rates for the PrPf in IndE (see above). Figure 5.10.16 shows a particularly high frequency for speech, where the value for the ICE-IND sample (111) is the highest across the board. The written genres, in contrast, are less PrPf-friendly. The latter result, too, tallies with findings from an earlier corpus-based register study (Balasubramanian 2009: 92–93, 157).

The dispersion of the normalized PrPf frequencies is above average ($\sigma = 33.10$) and the text types with the most extreme values are creative (w2f; 31) and persuasive (w2e; 140) writing. The strongest correlation of the normalized frequencies is in evidence with IrE ($r = 0.87$), although it has to be noted that a number of further correlations with other more PrPf-friendly L1 varieties, such as BrE ($r = 0.85$) and AusE ($r = 0.84$), are also strong.

The cluster dendrogram ($c = 0.56$) presented in Figure 5.10.17 shows a mixed result with regard to the associations between text categories. Three out of the four spoken types appear toward the right of the graphical representation but scripted monologues (s2b) seems to represent an outlier.

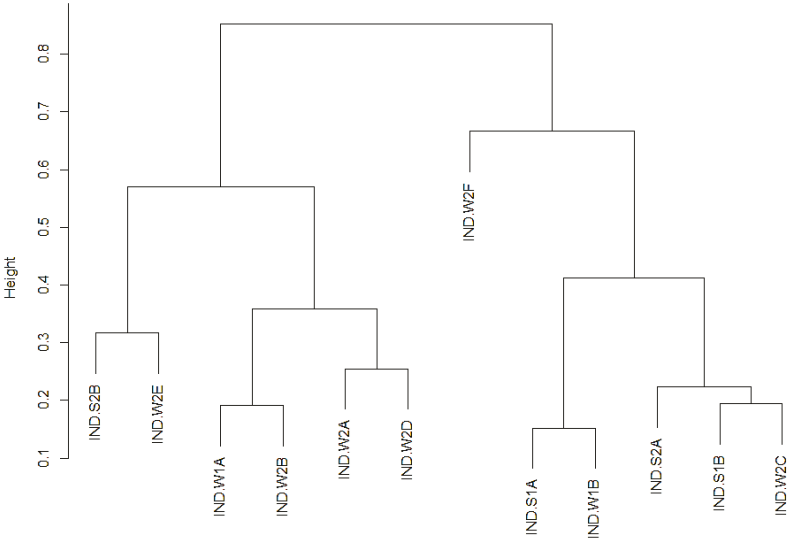


Figure 5.10.17. Cluster dendrogram of similarity across ICE text categories in IndE

Scripted monologues (s2b) aside, the partial alignment between the other spoken text types is illustrated in the non-hierarchical Neighbor-Net representation, which is shown as Figure 5.10.18.

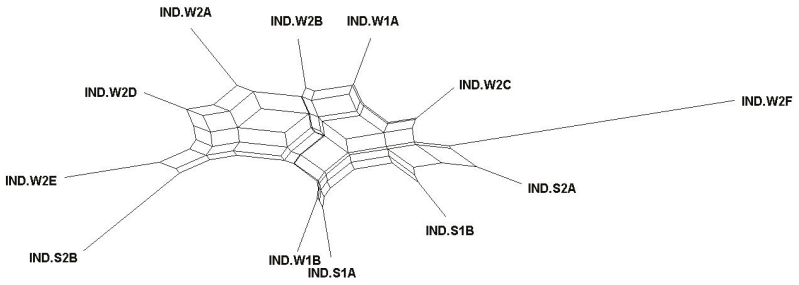


Figure 5.10.18. NeighborNet of similarity across ICE text categories in IndE

They cover neighboring nodes or are at least not distant from each other. A post-hoc test applied to the dendrogram in Figure 5.10.17 above shows further that public dialogues (s1b) and unscripted monologues (s2a) form a significant cluster. The conclusiveness of this result is weakened by the fact that the test also identified a significant cluster that contains all written text categories apart from creative writing (w2f) and letters (w1b), one of the involved written text categories. The latter once again appears close to the spoken types in Figure 5.10.18. Further clear outliers are the informational categories persuasive (w2e) and, most markedly, creative (w2f) writing. All things considered, we can conclude that the ICE-IND data sample, too, yields relative homogeneity of the spoken text types with regard to distributions of various factors determining the PrPf. Written texts, in contrast, show more variability, a finding that could be expected following Biber & Conrad (2009: 261).

6 The broader perspective

This section of the study presents a general approach toward assessing the divergence or overlap of varieties of English by way of comparing the various internal and external variables that were coded in occurrences of the PrPf. Therefore, it complements the preceding analyses by providing a bird's eye perspective on the data and abstracts away from a variety of singular observations. Rather than repeating each step of the individual analyses of the previous sections (temporal adverbials, Aktionsart, etc.) with the aggregated data from the ten World Englishes, it primarily relies on cluster and phylogenetic analysis. This is mainly due to the fact that these two methods have proven particularly apt for establishing latent structure in large sets of data such as the present one (see Chapter 4). In addition to an overall comparison of varieties, the structure of the ICE data further allows us to establish register, genre and text type effects, which is the second key aspect. This should help us to better tackle the following issues:

- Does the PrPf represent a core feature of English or are the patterns of usage (assessed by the language-internal factors) variable across varieties?
- Do theoretical notions such as areoversals or varioversals (adapted from Szmrecsanyi & Kortmann 2009) or variety types (according to Kachru's circles or Schneider's 2007 model) play a part in the variation?
- Or is variation determined rather by further structural factors such as register, genre or text types?

Another dimension that emerged in the course of the previous analyses and that may be of importance for the overall picture is the contrast between involved (dialogues, monologues, non-printed writing) and informational texts (printed writing) in the sense of Biber (1988).

The chapter concludes with three case studies (i) on surface forms that can be used as alternatives to the PrPf and their semantic and

pragmatic properties, (ii) on the combinability of temporal adverbials with the PrPf and the SPst and (iii) on evidence in ICE for usage of the PrPf as a tense that claims territory from the SPst. A discussion of the findings will follow in Chapter 7.

6.1 Overall comparison of varieties

The first point to investigate is the overall similarity of the varieties contained in our data. Based on the outcome of a hierarchical cluster analysis and of the NeighborNet view, it should become evident whether varieties (or groups of varieties) show the same intra-linguistic patterns as regards PrPf usage. The dendrogram (c = 0.85) presented in Figure 6.1 is based on a matrix with 10 x 312 cells established from the number of varieties multiplied by the number of individual factors such as type of temporal adverbial, Aktionsart of the main verb, preceding tense, etc. (26 values for each of the 12 text categories recognized in ICE).

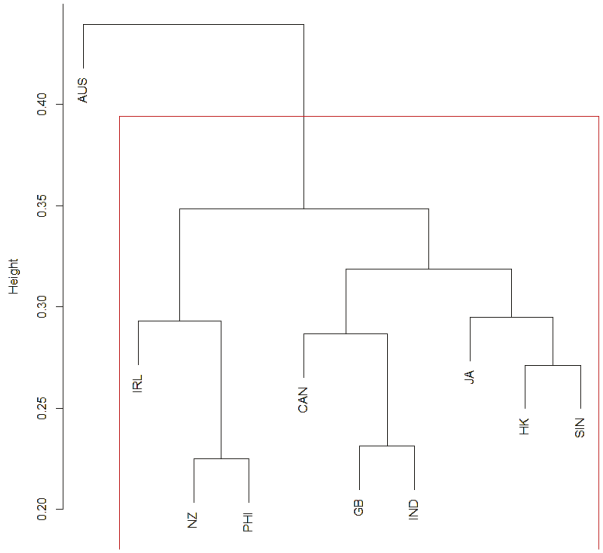


Figure 6.1. Cluster dendrogram of similarity across ICE components (language-internal factors)

Figure 6.1 provides a number of insights. First of all, it is apparent that neat clusters according to established variety type categorizations, for example L1 varieties such as BrE, CanE and AusE clustering together separately from L2 varieties such as PhiE, SinE and HKE, cannot be found in the present data. This seems surprising at first sight, but is very much in line with the findings of other corpus-based studies that cover domains such as the progressive (Hundt & Vogel 2011: 153) or progressive passives (Hundt 2009). Second, AusE is a notable outlier, apparently differing from the remaining varieties, which cluster closely in the highlighted area. Note, however, that the absolute height value even for the most dissimilar variety, AusE, is low (just exceeding 0.40), which is a first indication of the relative overall homogeneity of the varieties.[270] The large cluster includes varieties of all types and there is no evidence of a special status for transplanted L1 (phase 5) varieties, such as NZE, CanE or IrE, as they freely cluster with L2 varieties. Nor can geographical signals be observed. It is noteworthy that BrE forms a subcluster with IndE, the L2 variety that arguably still possesses a considerable degree of exonormative orientation toward BrE and that is considered to be in phase 3 of Schneider's (2007) phase model of Englishes (see Section 5.10). The phase 5 varieties can be found both in combination with each other (NZE/IrE) and clustering with L2 varieties (NZE/PhiE/IrE; CanE/IndE/BrE), while the rightmost cluster consists of L2 varieties that can be assigned to both phase 3 (HKE) and 4 (SinE, JamE) in Schneider's terms. This suggests that patterns of usage are not sensitive to variety type, or that this effect is only marginal and clearly one of tendency rather than an absolute split when varieties are considered globally.

As previous sections demonstrated the value of complementing hierarchical dendrograms with non-hierarchical representations, I will proceed accordingly.

[270] The relative homogeneity of the varieties could be confirmed by a multiscale bootstrap resampling approach in R, which identifies a single significant cluster (NZ/PHI) at a lower level only.

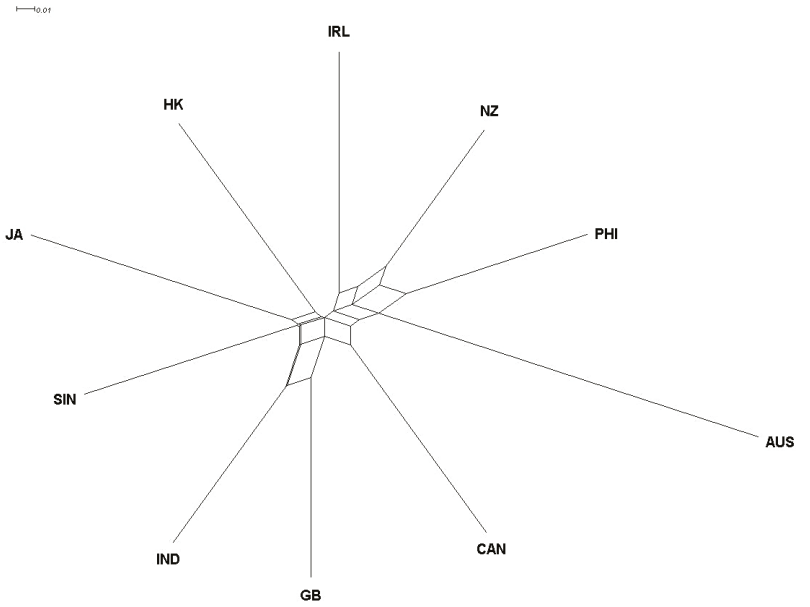


Figure 6.2. NeighborNet of similarity across ICE components (language-internal factors)

Figure 6.2 reveals that the importance of variety types might be underrepresented in the previous dendrogram. Generally speaking, the L1 varieties all group toward the right side of the network representation. A number of the apparent splits deserve closer inspection. First, one grouping that emerges comprises transplanted L1s (IrE and the Antipodean varieties) and PhiE. The remaining transplanted L1, CanE, is also close, but first shares a split with BrE. The L2 varieties align with BrE, while the data show that SinE and JamE, varieties that have further developed along the evolutionary circle (phase 4), share a split before they merge with the other L2s and BrE, the colonial ancestor. However, what is most striking is the star-like shape of Figure 6.2, which indicates that all varieties (with the possible minor exception of AusE) are approximately equidistant from one another. In other words, they differ by approximately the same amount, while having many shared characteristics.

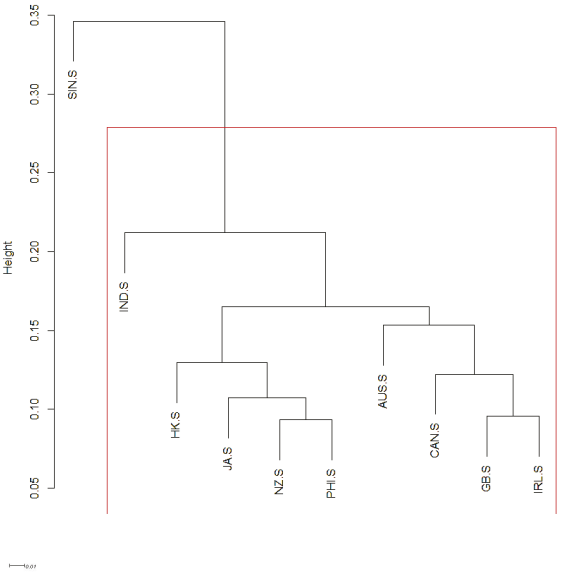
There are three possible explanations for this outcome. The relative homogeneity in distances between the varieties may be due either to

(i) the compilation principles of the ICE components, or (ii) a persistent strong exonormative influence of BrE, or (iii) the status of the PrPf as a core feature of the grammar of all varieties of English. While (ii) can most likely be rejected as a universally valid justification, as the outlier, AusE, is of a different variety type and some of the varieties have been oriented toward AmE (notably, PhiE, CanE and arguably JamE; cf. Sand 1999; see also Section 6.5), (iii), in interaction with (i), appears to provide the most probable account. ICE includes data from speakers who have at least finished formal English-medium secondary-school education, and are considered speakers of "'educated' or 'standard' English" (Greenbaum 1996b: 6). Within this type of educated English, which can be found across varieties of all types and locations, the patterns of usage for the PrPf are very similar when viewed globally (with some minor exceptions). Therefore, it seems legitimate to postulate the interim synopsis that the PrPf represents a globalized or core feature[271] of the grammar(s) of Englishes worldwide, although it has to be kept in mind that a different picture may emerge when social and dialectal variation or generic perfect contexts (allowing for other surface forms; see Section 6.4) are taken into account (see also Winford 1993; Miller 2000; Davydova 2011). The interpretation of the PrPf as a variety-independent core feature, however, may also have to be qualified after a more fine-grained analysis as in section 6.2.

6.2 Register, genre and text type effects

As has been shown throughout the preceding sections, the design of the ICE data enables us to track the effects of language-external categorizations such as register, macro-genre or text type on the distribution of individual factors. The same applies for the global analysis of the PrPf according to these dimensions.

[271] This definition of core is markedly different from that provided by van Rooy (2009). He tries to establish the "shared core" of the PrPf across a selection of L1 and L2 varieties and concludes that the most prototypical meanings of the PrPf are temporal (e.g. sequence of two events, non-continuativeness) while elaborations on these central meanings are more aspectual (e.g. terminativeness) in character.



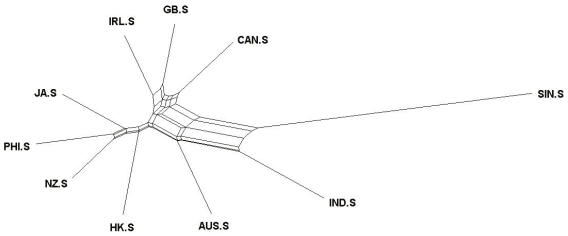


Figure 6.3. Cluster dendrogram and NeighborNet of similarity across ICE components (language-internal factors): speech

The dendrogram presented in the top half of Figure 6.3, which shows similarities across the spoken parts of the corpus components, is based on a matrix with 10 x 104 cells established from the number of varieties multiplied by the number of individual factors (26 values for each of the four spoken text categories recognized in ICE). It reaches a very high c-value (0.98), which indicates that the graphical output represents the actual data structure as contained in the matrix almost perfectly.

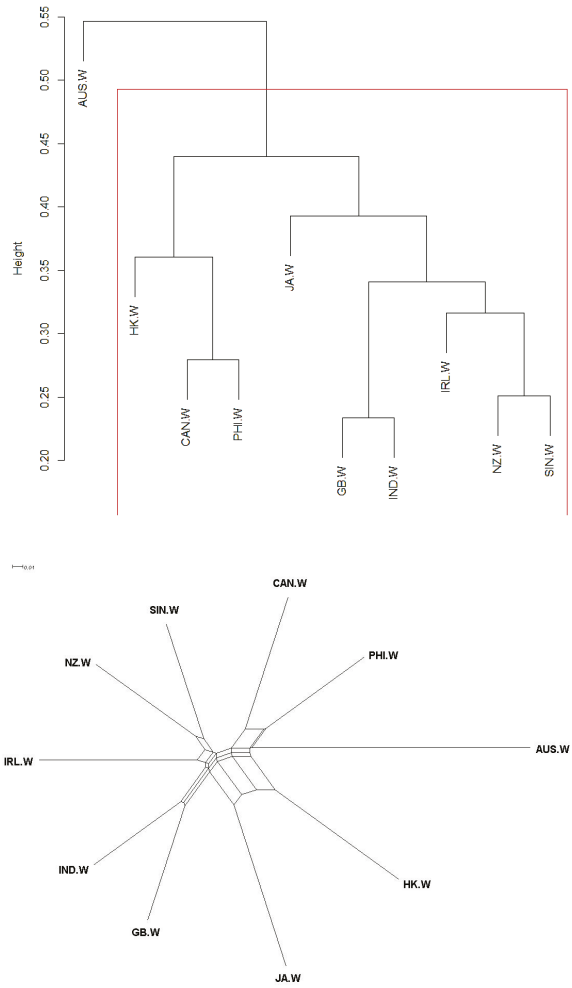


Figure 6.4. Cluster dendrogram and NeighborNet of similarity across ICE components (language-internal factors): writing

The top part of Figure 6.4 shows the dendrogram (c = 0.86) for the written counterpart. It is based on a matrix with 10 x 208 cells established from the number of varieties multiplied by the number of individual factors (26 values for each of the eight written text categories recognized in ICE).

A comparison of the two registers provides a number of insights. First, it is evident from figures 6.3 and 6.4 that neither a clear-cut alignment of varieties of the same type nor geographical signals can be observed (except for the rightmost cluster in Figure 6.3 that contains only L1s). Second, the above finding of a relative homogeneity across the varieties naturally also applies from the perspective of register.

In the dendrogram of the spoken parts (Figure 6.3), ICE-SIN constitutes the only outlier, while the remaining varieties form a close cluster (as highlighted in the figure). A post-hoc test reveals that only the JA/NZ/PHI subcluster is significantly different from the rest. Correspondingly, the NeighborNet shows that all varieties apart from SinE and, less clearly, IndE, group around an imagined center with only a few split signals, while the JA/NZ/PHI triplet is slightly displaced to the left. A similar picture emerges for the written part (Figure 6.4). Here, the AusE data surface as the only outlier in the hierarchical cluster dendrogram, while a post-hoc multiscale bootstrap test yields no significant clusters. The NeighborNet view, which is approaching a star-shaped figure, further validates the comparative uniformity of the differences between varieties in this mode of discourse as well. Groupings in terms of neighboring varieties of the same type or geographical location are not evident.

The next step in the analysis is to consider potential macro-genre effects; that is, to test whether the macro-genre categories group together across varieties. The related hierarchical dendrogram is shown in Figure 6.5 below.[272] Some groupings can be identified in there, at least in tendency. The leftmost cluster (3/3) and that labeled A (11/12) almost exclusively contain dialogues (DIA) and non-printed writing (NPW), which in many instances can also be seen as a sort of dialogic genre (as repeatedly exemplified in the analysis of individual varieties above). In contrast, nine out of the ten monologue (MON) leaves can be found in clus-

[272] The c-value of this representation is 0.77. Due to their differing sizes (two text types in dialogues, monologues and non-printed writing, six in printed writing), the comparison of the macro-genres had to rely on the aggregated relative values rather than on the detailed values for each text type contained. This in part also explains the comparatively low value of the coefficient.

ter B, which is otherwise dominated by printed writing (PW; 10/10), while dialogues (2/10) and non-printed writing (3/10) are rare.

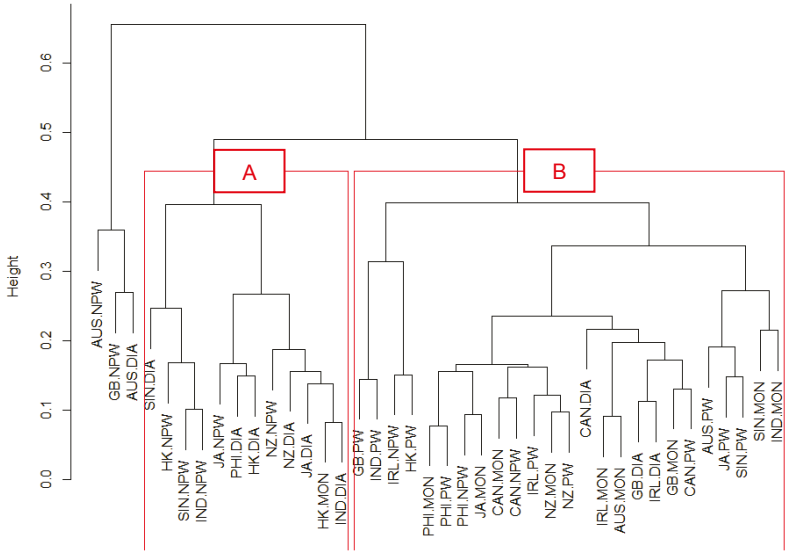


Figure 6.5. Cluster dendrogram across ICE components (language-internal factors): macro-genres

The association of the monologue and printed writing macro-genres is unexpected; thus, a NeighborNet representation (Figure 6.6) was computed to test whether this is an artifact of the hierarchical representation or rather a genuine alignment. Note in that respect that a multiscale bootstrap test for the significance of clusters in R returns a somewhat inconclusive result (not shown in Figure 6.5), with significant clusters that contain at least three different macro-genres. This would suggest that, overall, macro-genre effects are not manifest.

The NeighborNet in Figure 6.6 by and large confirms the results of the preceding hierarchical analysis. However, some tendencies with regard to the groupings can be identified. The majority (7/10) of the non-printed writing nodes can be found toward the left, sharing a number of splits. While most dialogues group toward the bottom of the representation, printed writing nodes are oriented toward the right. Thus,

printed writing can be viewed as the only macro-genre in which the individual nodes are closest to each other and thus most consistent (see also Figure 6.7, which shows a star-shaped NeighborNet that compares printed writing only). Monologues (with the exception of GB, IND and SIN) appear toward the bottom right. Also noteworthy in the bottom right corner is that macro-genre nodes of all four types are present.

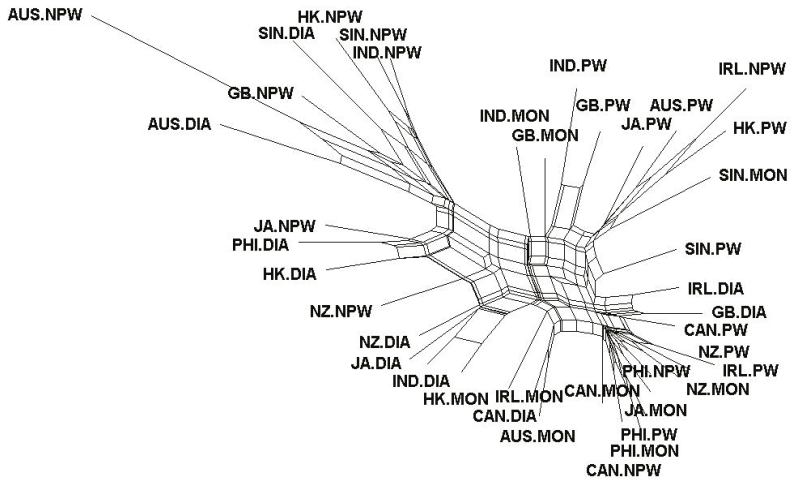


Figure 6.6. NeighborNet of similarity across ICE components (language-internal factors): macro-genres

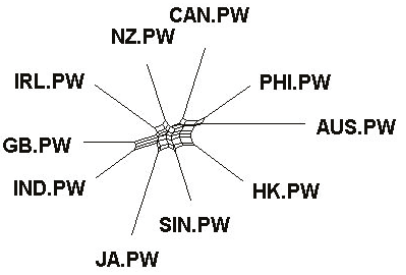


Figure 6.7. NeighborNet of similarity across ICE components (language-internal factors): printed writing

In sum, rather than a neat alignment of the same macro-genres from the different varieties, the NeighborNet (Figure 6.6) suggests a bipartite division between printed writing and the remaining categories. Whether this division, which could be relabeled "involved" versus "informational" in Biber's (1988) terms (cf. Section 4.1), can actually be verified is tested with regrouped data in a separate section below (see Section 6.3). One could also argue that, as in the above analysis of the cluster dendrogram, we have a division between dialogues/non-printed writing and monologues/printed writing. What speaks against this interpretation is the fact that one quarter of the nodes from the dialogues/non-printed writing pairing can be found among the nodes of the monologues/printed writing grouping. Further, while monologues appear to be closely aligned with printed writing, the majority of the relevant nodes are also close to dialogue nodes, such as those for ICE-NZ/JA/IND/CAN/IRL/GB. The macro-genres other than printed writing are internally much less homogeneous, but there are also no observable areal effects or alignment according to variety types, as the following NeighborNets show:

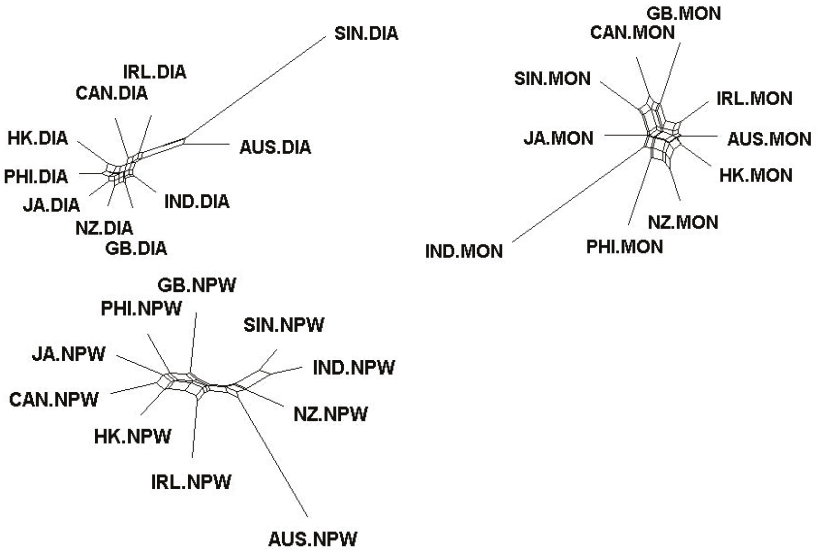


Figure 6.8. NeighborNets of similarity across ICE components (language-internal factors): dialogues/monologues/non-printed writing

As a final step, I now proceed to the analysis of potential effects of the twelve text types in the ICE data. This enables us to compare the similarity of the individual text types in each variety across the whole dataset. As with the analysis in the preceding section, this should clarify whether matching text types (or groups of text types) show the same intralinguistic patterns; in other words, whether distributional patterns are sensitive to the ICE text categories. The dendrogram (c = 0.85) presented in Figure 6.9 is based on a matrix comprising 120 x 26 cells established from the number of individual text categories (12 categories x 10 varieties) multiplied by the number of individual factors, such as type of temporal adverbial, Aktionsart of the main verb, preceding tense, etc.

As could be expected, variation across text categories, as indicated by the maximum height values where clusters merge, is more pronounced than across regional varieties, modes of discourse and macrogenres. In addition, the dendrogram yields a number of interesting observations.

First, it emerges that some texts of the same type are relatively homogeneous in terms of their distributions of the internal factors. That is, they appear in adjacent or even identical clusters. Above all, the vast majority (37/40) of the spoken texts (with AUS-s1a, SIN-s1a and IND-s2b representing outliers) can be found within a single cluster (labeled B in Figure 6.9). The homogeneity of this group does not come as a major surprise, especially after the analysis of the individual varieties, which showed that spoken texts in general align more closely than written ones (see sections 5.1 to 5.10 above). In addition, within cluster B spoken texts account for nearly three fifths of the leaves, such that this cluster can clearly be categorized as "spoken". Some further observations within the clusters at a subordinate level confirm this view. Cluster B1 almost exclusively (24/30) consists of spoken categories. Cluster B3 predominantly (8/11) contains private dialogues (s1a), while cluster B4 principally (15/19) comprises the spoken categories public dialogues (s1b) and monologues (s2a and s2b).

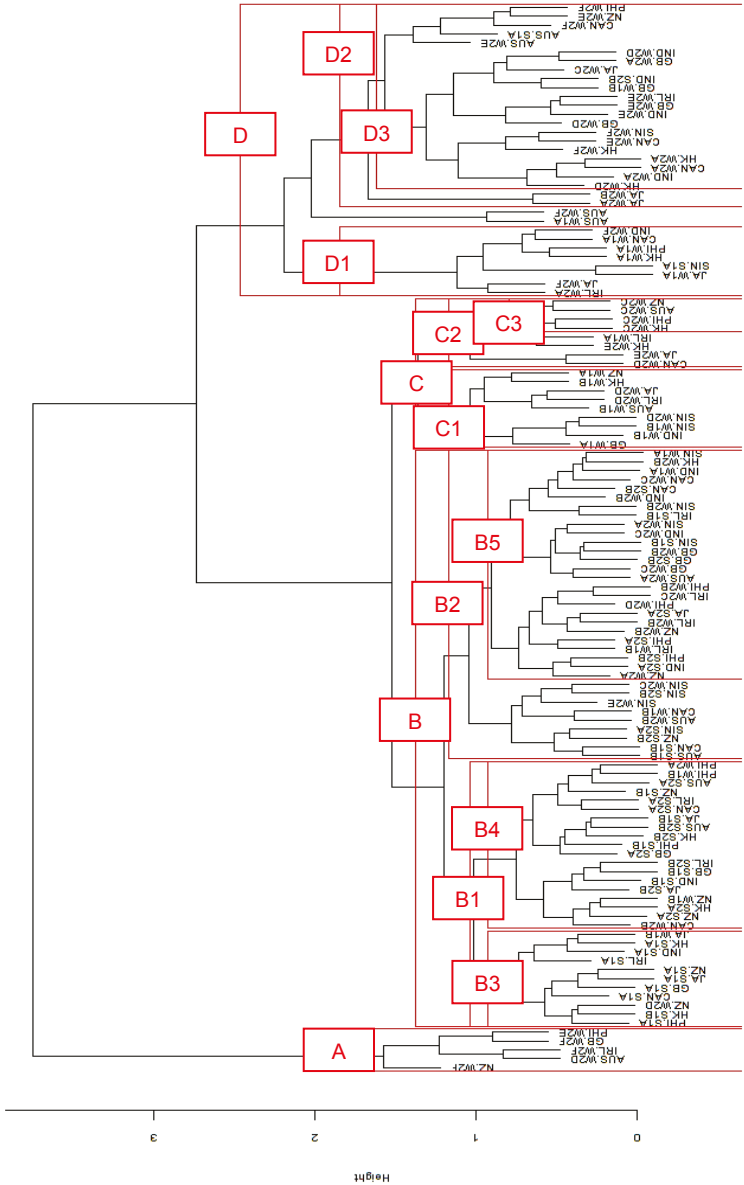


Figure 6.9. Cluster dendrogram of similarity across ICE text types (language-internal factors)

As mentioned above, the remaining spoken texts (apart from the three outliers) are all contained in cluster B2, although it must be conceded that spoken texts are less dominant in this cluster, with the distribution between spoken and written texts (13/35) nearing the default distribution of ICE (1/3 spoken versus 2/3 written).[273] Note that the distribution of the leaves is clearly not sensitive to variety type.

Written texts are to be found exclusively in cluster A and concentrate in the clusters toward the right of the dendrogram. Again, the members of some text categories can be found in the same clusters, while others are distributed more randomly. To be precise, cluster B5 comprises 7/10 varieties for popular writing (w2b), cluster D comprises 7/10 for creative writing (w2f) and cluster D3 comprises 6/10 for persuasive writing (w2e). Further trends are in evidence for the grouping of instructional writing (w2d) with 4/10 in cluster C, letters (w1b) with 4/10 in cluster C1 and academic writing (w2a) with 5/10 in cluster D2. Split clusters emerge for (i) student writing (w1a) with 5/10 in cluster D and 3/10 in cluster C, (ii) creative writing (w2f) with 3/10 in cluster A and, as mentioned above, the remaining leaves in cluster D, and (iii) reportage (w2c) with 4/10 in cluster B5 and 4/10 closely grouped in cluster C3. Like the findings for the spoken texts, the distribution of the leaves in the dendrogram does not seem to be susceptible to variety or variety type effects.

In sum, we can see that in the hierarchical cluster analysis the grouping of the various text types at different height levels is not random. This is especially true for the spoken categories, which are closely aligned. The picture becomes even clearer once we apply the broader ICE categories. Dialogues (14/20) and, less obviously, monologues (10/20) cluster in B1, while more than half of the texts included in non-printed writing agglomerate in clusters C and D1 (11/20). Printed writing, which comprises a wide range of different text types, is more diverse

[273] The proportions are reverse when taking into account the word count for the spoken (3/5 of the data; c. 600,000 words) versus written (2/5 of the data; c. 400,000 words) sections for each ICE component. However, this has no direct implications for the cluster analysis based on the given text categories.

overall but dominates in B2 (18), C2 (7), D2 (20) and A (5; overall: 50/60).

In addition, if we change the perspective to informational (printed writing) versus involved (the remaining text types) in the sense of Biber (1988), it is worth noting that 54 out of the 60 leaves that can be counted as involved are located in the aggregated cluster B+C, pointing to a potential division of the data along these lines (see also the next section). However, a test for the statistical significance of the clusters in R returns only a few clusters on a lower level. No clear division (e.g. with a significantly different cluster B, the one labeled "spoken" above) can be found, which in turn suggests that – with the exception of a few outliers – the data are homogeneous even from a more fine-grained perspective, although some (mostly non-significant) groupings can be observed.

The non-hierarchical view presented in Figure 6.10 below complements the previous analysis and offers some additional insights. Above all, in this graphical representation, too, no geographical signals or groupings according to variety types can be seen. While in the left third of the intricate NeighborNet the nodes are close (despite the presence of many splits due to the extensive number of taxa/nodes), as a general trend, the further we go to the right, the more distinctive the nodes become and a number of outlier groups can be established. First, creative writing (w2f) as a whole (10/10) seems to be distant from the imagined center, while its members are also comparatively dissimilar from one another. On the right, other, smaller groups which are conspicuous are persuasive (w2e; 6/10), academic (w2a; 6/10) and, less clearly, student (w1a; 5/10) writing. Second, categories with intermediate distances to the imagined center are instructional writing (w2d; 8/10 in two groups of four at the top and the bottom) and, again less clearly, letters (w1b; 5/10). Third, private dialogues (s1a; 7/10) assemble at the bottom and unscripted monologues (s2a; 7/10) at the top of the left third. Further, in relation to the spoken cluster identified above, it is worth noting that the vast majority of all spoken text types (37/40) can be found in this same area. At first sight, this indicates a clear split along the lines spoken (left) versus written (right). In fact, all nodes from popular writing (w2b) and reportage (w2c) are also located in the left third

semicircle; thus, this seems to be a split that is relative rather than absolute, or at least one that comes with certain qualifications.

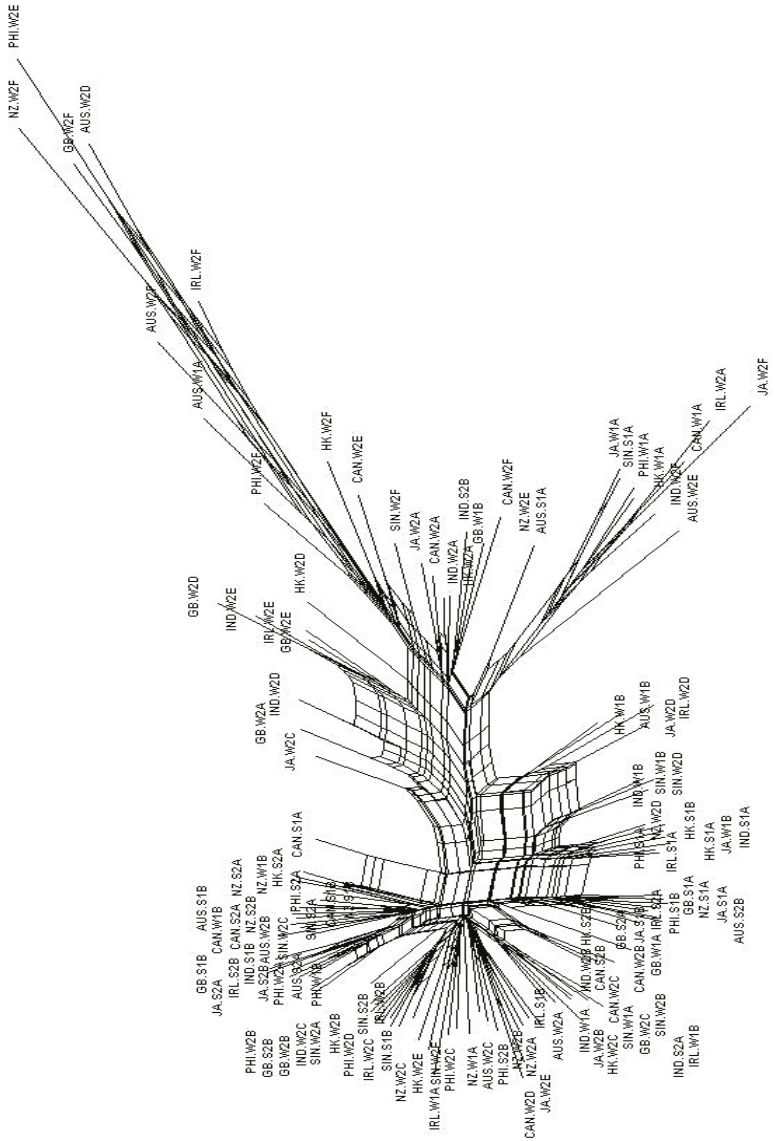


Figure 6.10. NeighborNet of similarity across ICE text categories

6.3 Involved versus informational texts

As indicated repeatedly in the previous sections, the findings by and large support the hypothesis of the PrPf as a core feature of English. Furthermore, in the course of the investigation it became clear that, rather than considering as dimensions of variation only those distinctions provided by ICE (that is, modes of discourse, macro-genres and text types), alternative divisions such as Biber's (1988) "Dimension 1" ("involved" vs. "informational"; see sections 4.1 and 5.1.4 above for the attribution of ICE macro-genres to either category) may have an effect as well (see also Xiao 2009: 441). Whether the reorganization of the corpus data according to this distinction yields any meaningful results is tested below.

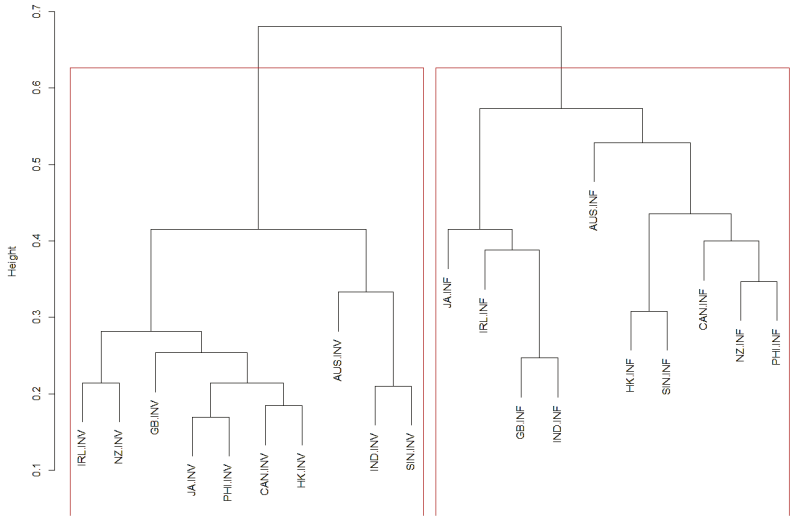


Figure 6.11. Cluster dendrogram across ICE components (language-internal factors): involved vs. informational

The cluster dendrogram (c = 0.78), as presented in Figure 6.11, shows a neat two-way grouping, as highlighted. All involved leaves can be found in the left cluster while all informational leaves can be found on the right. On the basis of the dendrogram, therefore, we can tentatively es-

tablish a meaningful division along the dimension involved versus informational.

In the NeighborNet (see Figure 6.12), all involved nodes group toward the left while all informational nodes can be found toward the right. The seemingly elegant picture is somewhat tarnished by the fact that some pairs of individual nodes from different categories are closer to each other than to the remaining members of their respective group (as is the case, for instance, for the involved node of AusE and the informational node of CanE). Note in this respect that a significance test in R for the clusters in the above dendrogram (Figure 6.11) indeed reveals two large significant clusters along the general lines involved versus informational, but that these clusters also contain some "foreign matter" in the guise of leaves that belong to the other category than the remaining leaves (i.e. clusters that again are significantly different from the remaining ones, e.g. the involved nodes of AusE, SinE and IndE), which ties in with the results from the NeighborNet. Still, the non-hierarchical representation largely confirms the finding from the cluster dendrogram above.

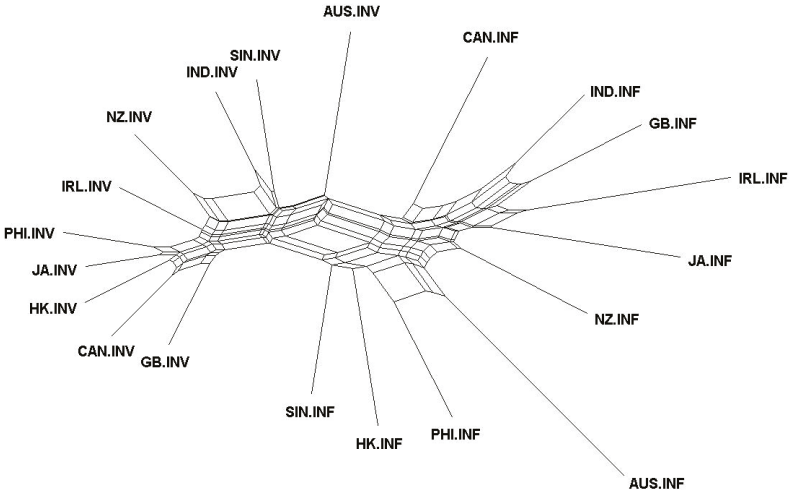
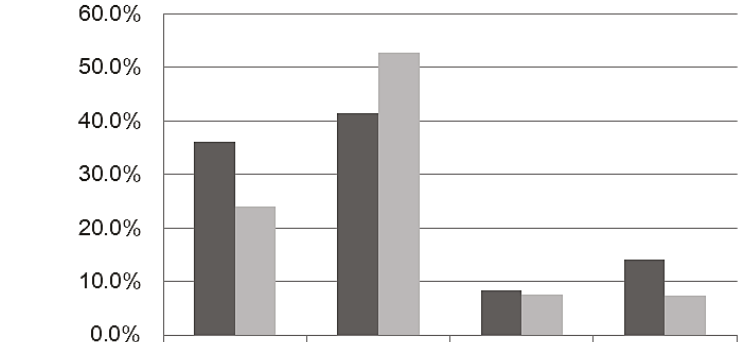


Figure 6.12. NeighborNet of similarity across ICE components (language-internal factors): involved vs. informational

The difference between involved and informational texts cannot be reduced to one particular variable. It is rather grounded in different distributions for individual factors or groups thereof. First of all, note that the differences for all groups are significant, except with respect to the presence or absence of temporal specification. However, this finding has to be qualified due to the large size of the samples, so that the differences amount to weak effects (φ_c = 0.065 or less) only. The only group that emerges with a meaningful value is type of temporal specification (χ^2 = 21.412, df = 3, p < 0.001, φ_c = 0.127). In detail, the values for specification for time-position (TP) are higher by one third and for adverbials of sequence (SQ) in fact twice as high in the involved texts, while specification for span and duration (SP) is less prominent (see Figure 6.13).[274]



	TP	SP	FR	SQ
involved	36.1%	41.4%	8.4%	14.1%
informational	24.1%	52.7%	7.6%	7.3%

Figure 6.13. Types of PrPf adverbial specification: involved vs. informational texts

[274] A chi-square test for significance with given probabilities (*chisq.test* function in R; probabilities are 7/10 for involved and 3/10 for informational texts according to the word count of the respective categories) shows that individual differences in absolute counts of TP (χ^2 = 19.5243, df = 1, p < 0.001) and SQ (χ^2 = 18.786, df = 1, p < 0.001) are highly significant, while those in SP are not.

Furthermore, there are noticeable differences in the relative values,[275] which thus contribute to the structural contrast in the following individual variables compared to other groups:

(i) for Aktionsart: activity verbs, which are more prominent in informational texts (involved: 36.3%; informational: 40.5%; differences are not significant);
(ii) for semantic readings: indefinite single acts, which are more frequent in the involved domain (involved: 64.7%; informational: 58.3%; $\chi^2 = 25.5695$, df = 1, $p < 0.001$); and
(iii) PrTs and SPst as preceding TR forms; the former occurs more frequently in involved texts (involved: 60.2%; informational: 56.0%; $\chi^2 = 25.7218$, df = 1, $p < 0.001$).

The salience of PrTs is arguably due to the more present-oriented discourse of involved texts, while SPst occurs more frequently in informational material (involved: 13.3%; informational: 21.0%; $\chi^2 = 13.4966$, df = 1, $p < 0.001$), arguably testifying to the stronger orientation of informational texts toward the past.

6.4 Case study 1: Alternative perfect forms

This section provides a qualitative analysis of a selection of alternative surface forms[276] that occur in perfect contexts in different varieties of English. Variation with the SPst or the PrTs as alternative forms or leveling between the PrPf and the SPst as occurring in L2 varieties (Kortmann & Szmrecsanyi 2004: 1188) will not be discussed here (but see sections 6.5 and 6.6) as a range of detailed analyses have already been conducted (see e.g. van Rooy 2009; Davydova 2011; Davydova et al. 2011 or Seoane & Suárez-Gómez 2013 for quantitative comparative corpus-

[275] For purposes of illustration, four categories with the largest percentage differences are shown. For a complete list see Appendix B.
[276] See also Section 3.4 above. For an overview of areal distributional patterns of these morphosyntactic forms and further tense and aspect forms, see Lunkenheimer (2012). Note that progressive and passive PrPf forms are included in this section in order to obtain a broader range of data.

based studies and Agnihotri 1988: 93; Winford 1993: 161; Davydova 2008; Balasubramian 2009: 157; Gut 2009: 273–274; Sharma 2009: 191 or Hundt 2013: 194 for remarks on individual varieties).

These alternative forms represent an interesting case, given that "variants that are regionally or socially marked are usually not maintained" (Mesthrie 1993; cited in Schreier 2010: 456) and are therefore even more conspicuous in educated language as contained in ICE. Findings from earlier literature are weighed against the present results, with a particular focus on the alleged idiosyncratic semantic and pragmatic values of these non-standard surface forms. A disclaimer applies here: The interpretation of individual examples is not always straightforward due to the strong presence of ambiguous and elliptical syntactic structures and to the potential presence of both performance and transcription errors characteristic of spoken data (see the discussion in Section 4.5). Note that from a quantitative perspective, variants are comparatively rare and occur predominantly in speech, but are more frequent in L2 varieties (Davydova 2011). Moreover, they have been studied widely as "trademark features" of IrE, which might also have bequeathed these to L2 varieties or other L1 varieties (such as CanE, as illustrated in example (1); see also Halford 2002: 42; McCafferty 2005: 242; Kortmann & Schneider 2006) in the process of colonization.

(1) "They*'re after changing* the music," says 79-year-old Joe Kennedy, as he takes a break from playing me some tunes in his cluttered house deep in the woods near Inverside. (ICE-CAN w2b-008)

This does not imply, however, that their study is irrelevant, as the full picture of a grammatical domain (modes of expression of perfect contexts) can only emerge if as many surface realizations as possible are included (see Kretzschmar 2012). This analysis focuses in particular on whether these non-standard forms can be seen as a generic substitute for the standard PrPf in all contexts, or whether they carry particular connotations and are thus restricted to specific domains.

The first type of non-standard surface form appearing in a present perfect context is auxiliary deletion of the type Ø + *been* + V-*en* (n = 102),[277] as illustrated in (2) to (10):

(2) Stills Ø *been linked up* to a television programme which was Understanding Northern Ireland and uh grabs were made just clicking on the screen and saying I want to save that image (ICE-IRL s2a-051)

(3) Police Ø *been asked* Inspector Heung at least being asked many many times about that (ICE-HK s2a-065)[278]

(4) Everything is set up like the uh the stove Ø *been built* and the uh air conditioning uh at that point at that (ICE-HK s1b-072)

(5) An uncomfirmed (sic) report that one motorist Ø *been arrested* on highway Thirteen south in that reserve not reserved lane (ICE-CAN s2b-006)

(6) So why why then would you see pictures Ø *been steamrolled* whereas I see as somebody who uhm paints something in simplistic terms a specific Rembrandt is just time and space (ICE-GB s1b-008)

(7) Since they're exceptionally gifted uh professionals Ø *been given* a good course of education there is some reluctance in a part of them to go to rural areas (ICE-IND s1b-041)

[277] The search syntax presented in Section 4.2 above is modified accordingly (e.g. search for *_VBN + *_VDN/*_VHN/*_VVN within four words to the right under exclusion of the *_VHB *_VHZ tags and subsequent deletion of irrelevant tokens). For reasons of space, selected examples are shown here and below, while the absolute number of occurrences for each alternative form is provided. Note that also perfects with bare past participles, such as *I Ø never seen him before* (ICE-CAN w2f-004) or *I Ø never seen anyone study like that Velma* (ICE-CAN w2f-004) and combinations of non-standard features, as in *So you been do something about* (ICE-HK s1b-073), occur sporadically in the ICE data. Further examples from NZE are shown in Section 5.4. Debatably, those could also be interpreted as leveling between past participles and SPst forms (eWAVE Feature 131 2012).

[278] See also the discussion of *been* versus *being* in HKE in Hundt (2009: 294). Cf. examples (6) and (25) below, which may be seen as instances where performance or transcription errors complicate an unambiguous interpretation.

It is evident from the examples that this elliptical form appears in both L1 and L2 varieties and naturally occurs in passive constructions. Note, furthermore, that in examples (6), (7) and (10) below the preceding relative pronouns are also omitted.

Nevertheless, the form is not restricted to passives. This is illustrated by examples (8), where *been* is used not as an auxiliary indicating the passive voice but as a full verb (see also (9) with the spelling variant *bin*),[279] and (10), where pre-verbal *been* is used instead of the usual auxiliary *have*.

(8) i said oh you know where Ø you fellas *been* (ICE-NZ s1a-087)
(9) Poor Tweetie, he says, you Ø *bin* through somethin'? (ICE-NZ w2f-011)
(10) Uhm you know in recent times we have heard of a number of demonstrations Ø *been taken* place at our schools disrupting uhm the the operation of schools and the performance of children (ICE-JA s1a-096)

The last example is particularly conspicuous as the statement contains variation between the PrPf and the non-standard form, and *been taken* arguably shows traces of a tense form typically found in Jamaican Creole (*ben* + V; Patrick 1999: 171; Kortmann & Schneider 2006) that has become partly "standardized". (10) may alternatively be interpreted as elliptical progressive form (Ø *been taken* instead of Ø *been takin'*).

Further, it is worth noting that a number of nouns preceding the Ø + *been* + V-*en* construction end in a sibilant (mostly due to the presence of plural forms; but cf. *Police* in (3) and *motorist* in (5), where the sibilant is likely to occur in word-final position due to cluster reduction/deletion of /t/). It can be speculated that the presence of a sibilant before *been* provides an explanation for HAVE-ellipsis in connected speech, as it resembles and implies the cliticized third person form (*has* > *'s*). In analogy to the label "apparent preterite", by means of which Elsness (1997: 347) attributes the decreasing frequency of the PrPf to the

[279] Hundt et al. (2004: 583–584) have noted that HAVE-dropping is common in NZE with highly frequent verbs such as BE.

growing use of contracted auxiliary forms and resulting phonological congruence of PrPf and SPst forms, we could label this phenomenon the "apparent perfect". It goes without saying that this line of argumentation only holds under two assumptions: First, that the transcriptions reflect the grammatical structure intended by the speakers and, second, that some of the examples represent instances where speakers do not follow standard patterns as regards congruence between the plural noun and the auxiliary, as is further illustrated in (11) to (13) from spoken HKE, where this feature seems to be particularly salient in the data. In these examples, we may have to consider that the transcription in the corpus data leaves the issue of congruence and morphology open, as in (11). *Responsibilities/responsibility's been put* can be seen as homophonous variants and it is impossible to judge which of the two is the form actually intended. In fact, singular interpretations are more likely for *systems/system's been changed* in (12) and *works/work's been poured* in (13). However, such an account is less probable in the examples where the relative pronouns are also omitted (see above) and in those involving progressive forms (see (14) and below).

(11) I think a lot a lot of *responsibilities been put* on the RAs (ICE-HK s1b-077)
(12) I believe if the uh open door policy continue in China and the uhm the social *systems been changed* and the accounting standard or accounting rule must follow (ICE-HK s2a-023)
(13) Okay this is his first run in almost four weeks Seattle Sun so the *works been poured* into him (ICE-HK s2a-014)
(14) But um certainly there are these little changes *Ø been going on* in the way people use words (ICE-AUS s2a-050)

HAVE-ellipsis also repeatedly occurs with progressive forms (n = 28), as exemplified below:

(15) I always *Ø been playing* the role as a more like a a [unclear word] of assisting helping uh uh discussing policy and things like that (ICE-HK s1b-046)

(16) Yeah Ø *been howling* for like two or three hours (ICE-IRL s1a-094)
(17) Ø *Been doing* some more eden grogan ware (ICE-AUS s1a-074)
(18) Ø *Been doing* anything interesting there (ICE-GB s1a-095)
(19) They Ø *been delivering* beer to different restaurant hotel and pub (ICE-HK s1a-063)
(20) She said to me say uhm she see me stand up by the bus stop and she Ø *been looking* at me and its just like she says she have to say something to me so just (ICE-JA s1a-045)
(21) We would like to speak to you because we Ø *been getting* a lot of callers who are suggesting by implication that Mister Anderson might have had some unfair advantage in the system by virtue of being a Member of Parliament (ICE-JA s1b-050)
(22) Ø *Been eating* hellava lot lately, too much if you ask me. (ICE-SIN w1b-004)
(23) I Ø *been talking* to former employees um and employees a number of times that's correct (ICE-AUS s1b-070)

A look at the wider context in which the examples are embedded reveals that it appears across various possible semantic readings, such as indefinite (e.g. (17) or (18)) and continuative (e.g. (15) and the progressive forms), or more specific interpretations such as recentness (e.g. (21) or (22), where the recentness is made explicit by a temporal adverbial). Considering that it occurs in varieties of all types and exclusively in spoken registers (with the exception of (22), apparently taken from an informal personal letter) this non-standard surface realization can best be seen as a vernacular feature with no particular semantic implications. One finding with repercussions for the pragmatic value of this form is that it, instead of the PrPf, can be widely found in all stylistically appropriate (that is, mainly spoken informal) contexts. Naturally, whether this is seen as a proper pragmatic function depends on whether one subscribes to a broad or narrow conception of pragmatics.

In the ICE data, a number of forms with various realizations of auxiliary BE instead of HAVE also occur (n = 114):

(24) Yes she *is* been wanting to do that (ICE-SIN s1a-035)
(25) It's kind of vaguely related to this piece that I *was* been talking (ICE-HK s1a-019)
(26) I*'m* just come here on a holiday (ICE-IND s1a-001)
(27) So not only *are* the consumers' purchasing power been deteriorating so have their confidence (ICE-SIN s2b-038)
(28) the thing is that sometimes when they*'re* been out a lot during the week (ICE-NZ s1a-013)
(29) And he's come up with this check list of things he *is* given to his students (ICE-HK s1a-053)

While BE-perfects regularly surface in L1 dialects[280] and are common in older stages of English, in ICE these are mainly found in Asian L2 varieties. In principle, there may be two explanations for their occurrence.[281] On the one hand, assuming the validity of the "founder principle" (Mufwene 1996; cited in Hickey 2004), they could be seen as remnants in the Asian varieties of the colonizers' dialects. This seems unlikely as BE-perfects as a rule can be found with mutative verbs only (Shannon 1995: 147–150), and (26) is the only example of this type. On the other hand, substrate influence (see e.g. Davydova 2011: 206–208 with respect to Hindi influence on IndE) and learner effects may play a role here. In general, Asian speakers of English, mainly due to influence from substrate languages, use less verbal inflection for the expression of temporal and aspectual relationships and rely on other means, for example mark-

[280] See the relevant contributions in Kortmann et al. (2004). Cf. also Allerton, who describes the BE-perfect as "American conservatism or an immigrant Americanism" that "seems to be slipping in [to BrE; V.W.] almost unnoticed" (2008: 32–33).

[281] A simple transcription error due to phonological similarity of the reduced forms (/z/ vs. /(hə)z/) might be present in example (24), given that it represents transcribed (unscripted) dialogue. Cf. also *He* is *he* is *always been with DL you know* (ICE-SIN s1a-032) or *I mean he use* (sic) *to be a real a real cool guy so you know getting girls* is *never been a problem for him* [...] (ICE-SIN s1a-058). However, this is unlikely for the remaining examples, in particular in (29), where a cliticized form appears in the same utterance. Another area where inconsistencies in the transcription occur due to phonological similarity is *been* for *being* (or its colloquial realization /biːɪn/), e.g. *Okay once the pieces are* been *cut and washed and dried now we connect them together* (ICE-SIN s2a-058). On a related note, this may lead to non-standard uses in written production (in L1 varieties), e.g. *That is, the institutionalization of labour relations are* been *seen as progressive* (ICE-CAN w1a-007).

ing by adverbials or particles (see the examples in the respective sections of Chapter 5). For the present examples, which are in line with the standard PrPf except for the auxiliary, this can be excluded. However, it has been suggested for SinE that an internal continuum from a pidgin-like to a standardized variety exists (Tickoo 1996: 446), and it would appear safe to assume that this also applies for other Asian varieties (see also above). Therefore, it seems plausible that the examples illustrate a position somewhere in the middle of the continuum. This hypothesis is strengthened by the fact that all the examples are from spoken sections, where acrolectal/upper-mesolectal speakers, too, are more prone to using non-standard features. This is further illustrated by example (27), where another variant ([*their confidence*] *have been deteriorating*) approximating the PrPf (see also Bautista 2008: 217) occurs in the immediate context.

Regarding the pragmatic value of the BE-perfects, a similar assessment as in the preceding section can be given, the difference being that BE-perfects in the ICE data seem to be a feature of the spoken vernacular of L2 and in particular Asian varieties of English. However, as a larger set of (especially spoken) data may yield different results, these conclusions remain tentative.

Occasionally, the verb remains entirely unmarked:

(30) I just *come* in Hong Kong by myself (ICE-HK s1a-070)[282]
(31) The Minister *remind* the House a few moments ago that Changi airport is also number one in the world (ICE-SIN s1b-060)

A variant of the unmarked version emerges as a kind of intermediate or approximating pattern, nicely illustrating the continuum between standard-like and mesolectal/basilectal forms (see also Winford 1993: 150). Here, speakers use a marked form of the auxiliary in combination with a generic verb form (n = 113), as the following examples show:

[282] An alternative interpretation of this example would be auxiliary deletion of the type Ø + V-*en* (cf. above).

(32) The Ministry *has review* its practice on making a rigid or fine distinction between hospitalization and outpatient medical leave (ICE-SIN s1b-058)
(33) Two that the learned judge below *has exercise* his judicial discretion correctly in respect of the interest awarded and his order to stay proceedings pending arbitration of the sum of five hundred and nineteen two hundred and eighty-four (ICE-SIN s2a-064)
(34) I *have waste* so many money to to buy a ticket for that (ICE-HK s1a-081)
(35) But that uh [...] visa or something *has change* no (ICE-IND s1a-045)
(36) A very simple catch no fun fair the Barbados team just gathering together in the middle of the field but I think they *have recognize* the fact that they are so far ahead its in almost impossible (ICE-JA s2a-003)

As indicated above, not marking the verb for tense or aspect is a typical property of Asian Englishes that is most likely due to substrate influence. However, here it is not restricted to Asian data, or, more specifically, to data where a Sinitic substrate is likely, as in examples (32) to (34). Thus, it can be hypothesized that at least the variant involving a combination of auxiliary and unmarked verb form is also due to imperfect learning or should be seen as a feature of L2 varieties in general (cf. the presence of the feature in IndE and JamE), a finding that corresponds with the results of an aggregational study of L2 varieties of English (Kortmann & Szmrecsanyi 2009: 279). Furthermore, it should be noted that these forms are very rare in the current data. Again, no specific semantic value of this surface form is in evidence.

The peculiarities of surface forms in perfect contexts in IrE and their potential origin(s) have been widely discussed in the literature (e.g. by Kallen 1989; Filppula 1997, 2004; Siemund 2004; McCafferty 2005; Kirk & Kallen 2006, 2007; O'Keeffe & Amador Moreno 2009; Pietsch 2009; Winford 2009; Hickey 2010a; see also Roberts 2007: 405–406), where the following realizations are recognized (adapted from Filppula 2004: 74–76; cf. Siemund 2004: 403–406; Pietsch 2009: 529):

a) PrPf (HAVE + V-*en*)
b) BE-perfect – resultative perfect (intransitive verbs)
c) SPst (*Were you ever in Kenmare?*) – indefinite anterior perfect
d) Simple Present (*I'm in here for about four months*) – extended-now/continuative perfect
e) *after*-perfect – recentness/hot-news perfect
f) Medial object perfect (MOP) – resultative perfect

While a) can be used in all contexts and is the most frequent variant (Fritz 2006: 291), it has been proposed that the remaining variants constitute a self-contained complementary system with the semantic-pragmatic values as indicated (Pietsch 2009: 529). For the present analysis, variants e) (n = 7), as exemplified by (37), and f) (n = 45), as exemplified by (38),[283] both "relative newcomers in the history of IrE" (Fritz 2006: 294), are of particular interest as supposedly unique features of IrE (but cf. eWAVE Feature 97 2012; eWAVE Feature 98 2012).

(37) In the opening round I thought for a while that Walsh was going to win inside the distance but he*'s after running* into a couple of hard ones here from Barrett (ICE-IRL s2a-012)
(38) So she *has her schoolbag packed* with her pencil case and that and her bits and pieces that she'll never have out for the first six months you know (ICE-IRL s1a-001)

Apart from the abovementioned function of expressing recentness, the *after*-perfect is seen as a characteristic pragmatic means of IrE (i) to convey intimacy and solidarity among its speakers (Kallen 1989: 31), (ii) to foreground a minor event that is closely related to the main event in a narrative, (iii) to (mock-)scold oneself, and (iv) to create a "mirative effect", that is to mark surprise about a piece of news (O'Keeffe & Amador Moreno 2009: 523–529). It should be mentioned that function (iv) reveals an interesting parallel to AusE, where it is claimed that the stand-

[283] See also Kirk & Kallen (2007) and O'Keeffe & Amador Moreno (2009) for further IrE examples.

ard PrPf increasingly fulfills a similar mirative function in police news reports (Ritz 2010: 3410). Note, however, that the data from ICE do not allow for a pragmatic interpretation besides the implication of recentness (see also McCafferty 2005: 241, 245–248). As regards MOPs (HAVE + object NP + V-*en*), it is worth noting that comparable forms also occur in other varieties, as shown by the following ICE data:

(39) they have agreed uh they already *have a plan set up* where uh any charitable organization that buys a ticket is donated a couple of dollars uh from the Expos (ICE-CAN s1b-041)
(40) Craig Murray up at Lewis shoes *has the shoes recently ordered* for my approval. (ICE-CAN w1b-007)
(41) Pritchet has also worked on the Canada-France-Hawaii telescope in Hawaii and along with VandenBerg and Dr. David Hartwick, *has time booked* on the Hubble space telescope. (ICE-CAN w2b-021)
(42) Smith now *has his sights set* on the Olympics of course next summer in Altanta (sic) (ICE-CAN s2b-001)
(43) andrei agassi *has his sights set* on his first grand slam title after overpowering second seed boris becker (ICE-NZ s2b-010)
(44) er there are rumours that he *has money kept* in his hut but i don't i really wouldn't know if those are true or not (ICE-NZ s1b-065)
(45) bachop *has his backs lined out* to the right to this side of the field (ICE-NZ s2a-001)
(46) Professor Johannes Itten *has extensive documentation gathered* from his students demonstrating each student's own private conception of subjective colour harmony. (ICE-AUS w2d-014)
(47) And I'd swear that your Mum probably as a part of her masochistic nature *has her scales turned* a bit the wrong way (ICE-GB s1a-011)
(48) That is why the present government also *has this bill actually already enacted* into law so that there could be equal distribution of wealth (ICE-PHI s1b-004)

(49) These cash wealthy companies, instead of re-investing in their companies and in real production, all *have their agents lined up* at the Bank of Jamaica and Stock Brokers to invest in paper. (ICE-JA w2e-005)
(50) Many insurance companies *have their mainframe operations outsourced* to Teleglobe Insurance Systems (ICE-HK w2a-033)
(51) Anyway they *have a corner ball hit* and now let us see what uh uh this Ranjit Kumar can do (ICE-IND s2a-012)

While MOPs occur across all regional varieties and variety types (*pace* Fritz (2006: 294) also in AusE) and are well established in both spoken and written texts, they seem to be particularly salient in CanE and NZE. Hickey argues that "the retention [of the MOP; V.W.] in Irish English and the use of this word order to express a resultative perfective can in large part be accounted for by the wish of Irish learners of English to reach an equivalent to the category of resultative perfective which they had in their native language" (2010: 157; cf. Pietsch 2009: 530). As this explanation does not readily apply for the remaining varieties, influence from a variety of English of Irish settlers that left its mark in the process of colonization, as can be seen in the *after*-perfect (Halford 2002: 42), presents itself as a likely scenario for the presence of MOPs in some varieties. In addition, the presence of MOPs in other varieties (notably, BrE, JamE and IndE) can best be analyzed as persistence of the feature from older stages and/or dialects of BrE (Winford 2009: 212–214; Kallen 1989: 16), although the Irish settler hypothesis cannot be excluded altogether.

From a semantic-pragmatic point of view, the MOP (in IrE) has been viewed as a construction emphasizing the result of an action, and this interpretation indeed seems to be valid for the majority of the above examples. However, implications of recentness, as in (40) and possibly (51), as well as continuativeness, as in the examples from sports reportage ((42) and (43)), can also be observed. A causative interpretation seems possible in (50).

In brief, a split picture emerges from the analysis of instances of "Irish" perfects (*after*-perfects and MOPs) in ICE. *After*-perfects are in-

deed used in contexts where the focus is on the recentness of an event, while previous studies have identified a number of other semantic and pragmatic values that do not surface in the current dataset.[284] MOPs, in contrast, rather than representing forms with specific values, are not restricted to a particular semantic or pragmatic domain and may substitute the standard PrPf in a variety of contexts. This latter finding qualifies Fritz's statement that the two types of perfect in question "are only used when an author felt an urgent need to express a particular meaning" (2006: 291), and corroborates Kallen's conclusion that a neat one-to-one relationship between each IrE perfect surface form and a single potential semantic or pragmatic reading or interpretation (the so-called "grammaticalization hypothesis") is in fact not a linguistic reality (1989: 31; cf. McCafferty 2005).

Furthermore, the data revealed a number of negated perfect contexts involving *ain't* + V-*en* (n = 4):

(52) they *ain't made* the glass slipper yet that would fit your big hoof. (ICE-NZ w2f-013)
(53) you *ain't seen* nothing yet (ICE-NZ s1b-033)
(54) it *ain't seen* nothing in it yeah (ICE-NZ s1a-087)
(55) Other say you *ain't seen* nothing yet (ICE-SIN s1b-026)

In the ICE data, negativized auxiliary pre-position *ain't*, which has been identified either as an original feature of Southern American or Black dialects or as an informal expression (Algeo 2006: 21), occurs in SinE and NZE. The notion of informality is furthered by the co-occurrence of negative concord or by the lexical choice (e.g. *big hoof* 'large foot'). While substitution of HAVE with *ain't* is fairly common in NZE (Kortmann & Schneider 2006), the occurrence of this variant in SinE can most likely be explained by the mere fact that particularly *ain't seen nothing yet*, as

[284] Note that the *after*-perfect rightly can be categorized as a pragmatic marker of "Celticity" (Kirk & Kallen 2006: 98), even if it occasionally occurs in further varieties of English, such as CanE (see example (1) above). Thus, the *after*-perfect seems to represent an exception to Trudgill's hypothesis that "it should be the frequency of features alone that accounts for adoption" (cited in Schreier 2010: 456). Here, social factors are responsible for the persistence of a characteristic alternative surface form.

popularized by the homonymous song by Bachman-Turner Overdrive, has developed into a stylized fixed phrase that has gained worldwide currency. The findings of this case study are further discussed and contextualized in Section 7.1 below.

6.5 Case study 2: Temporal adverbials

Although the current analysis has for the most part considered the PrPf on its own and, unlike many previous studies (e.g. Hundt & Biewer 2007; Yao & Collins 2012), not its alternation with the SPst, the current section serves as an exception to this general principle and presents a case study of one particular variable, temporal adverbials. Specifically, it explores co-occurrence patterns between a set of temporal adverbials and either the PrPf or the SPst.[285] This is motivated by the fact that specification via a temporal adverbial is commonly viewed as an important constraining factor (Jespersen 1931: 61) which may have ramifications for other aspects on various levels of linguistic analysis that are related to the PrPf (such as recentness of the event, semantics of the construction and pragmatic notions like the presence or absence of current relevance).

It might be argued that temporal adverbials are merely a side phenomenon as regards the variation between the PrPf and the SPst, as the analysis of individual varieties above has shown that the vast majority of PrPf occurrences remain temporally unspecified (see also Schlüter 2006: 143; but cf. Miller 2000: 350). However, as rightly noted by Hundt

[285] For this analysis, the full ICE dataset was used and 6,535 relevant occurrences were retrieved. Varieties are considered globally and comparability as regards registers, macrogenres and text types is thus negligible. The analysis also includes data from ICE-EA and a preliminary version of ICE-NIG, as kindly provided by Ulrike Gut (University of Münster). Note that only the written sections of ICE-NIG (c. 400,000 words) were available. To obtain comparable datasets, the number of occurrences of the temporal adverbials had to be extrapolated (multiplied by 2.5, as the spoken part of the ICE components usually comprises c. 600,000 words). This achieves the same effect as using normalized frequencies (of e.g. n occurrences per 10,000 words). It should be kept in mind, however, that all findings for ICE-NIG are of a preliminary nature. As all ICE components are of equal size, the dendrogram and NeighborNet below are based on absolute numbers. The search syntax is explicated in Werner (2013b), which also includes detailed analyses of individual adverbials.

& Smith "[t]his does not mean [...] that there is no mileage in investigating cases where the PP [=Present Perfect; V.W.] and SP [=Simple Past; V.W.] tenses are used with temporal adverbials" (2009: 52), as this type of indefinite adverbials nicely illustrates truly variable contexts between the PrPf and the SPst.[286]

As pointed out above (see Section 4.3.1), it has repeatedly been stated in the extant literature on the PrPf that, especially in (standard) BrE, severe restrictions apply as to the combinability of the PrPf with temporal adverbials that refer to a definite point in the past,[287] which render utterances like (1) ungrammatical or at least odd (but see Section 6.6).

(1) *We've been at the meeting yesterday.

This constraint on combinability is known as the "present perfect puzzle" (Klein 1992; cf. Portner 2003; Molsing 2006 or Rothstein 2007) due to its typological oddity in contrast to perfects in other European languages. Note, however, that this restriction does not apply for non-finite forms, as in (2) to (4), including forms that are modified by a preceding modal, as in (5).

(2) the process has now slowed down considerably, *having passed* its peak in the early 1970's (ICE-GB w1a-013)
(3) Michael Heseltine *having been* through all that ten days ago went to his campaign office to work on maintaining and increasing his first ballot vote (ICE-GB s2b-003)

[286] The need to identify a variable context has previously also been recognized by Wynne (2000). From a methodological point of view this is a non-trivial task, as true variable contexts can only be detected through an extended amount of manual analysis (see Davydova 2011: 119–124) and arguably subjective identification criteria. Note again that an older corpus-based study has found that temporal specification by adverbials is much more common with the PrPf than with other grammatical forms such as the Simple Present (c. 2% of all instances) or the SPst (c. 9%) respectively (Ota 1963: 43).

[287] These include relative adverbials (e.g. *yesterday, last week, two months ago*; Schopf 1984: 16) that refer to a point wholly in the past. For a general overview of the combinability of temporal adverbials in English, see Declerck (2006: 591–633).

(4) The taste for public monuments was considered *to have broken out* into a statue monument memorial mania in the nineteenth century (ICE-IRL s2b-031)
(5) The translucence of the petals *might have attracted* the eye about an hour ago. (ICE-SIN w2f-020)

On the other hand, there are temporal adverbials where this constraint does not apply. These characteristically refer to an indefinite or recent moment in the past or express (continuative) duration.[288] Examples include:

(6) Packing up has *already* begun (ICE-GB s2b-003)
(7) To-day is Wed & the passport hasn't turned up *yet* (ICE-IRL w1b-002)
(8) But her eyes, and the way she looks at me, they have *always* been that way (ICE-SIN w2f-019)
(9) and that was the longest party I have *ever* gone to (ICE-JA s1a-079)
(10) I want to say that we have *never* opposed Quebec's recognition as a distinct society (ICE-CAN w2e-003)
(11) And we have *recently* uh uh established that agrobactor (ICE-IND s1b-046)
(12) A woman came in from the kitchen and she said we're serving dinner in the dining room car now and we've *just* hit a two year boy (ICE-AUS s1a-036)
(13) This has *since* become a building block in public key infrastructure, ensuring the quality and legality of electronic transactions (ICE-HK w2b-035)

This group of adverbials (*already*, *yet*, *always*, *ever*, *never*, *recently*, *just*, *since*) presents an interesting case as its members have been conceived as more neutral given that, in principle, they allow for variation between

[288] See Biber et al. (1999: 467). Hundt & Smith subsume these adverbials under the label "adverbs of current relevance" (2009: 52), despite the problematic connotations of the latter term.

the PrPf and the SPst (Markus 1977: 57; McCoard 1978: 130; Sempere-Martinez 2008: 131). The variation is regionally determined, and while PrPf use with these adverbials has been shown to be categorical or at least more dominant in BrE, AmE has been claimed to be prone to the SPst or at least to allow for more variation (Quirk et al. 1985: 4.22; but cf. Panzner 1995: 51). That is, in AmE the relative frequency of the SPst in contrast to the PrPf is higher in instances that involve an indefinite temporal adverbial, particularly in less formal registers, hence the term "colloquial preterite" (Vanneck 1958; cited in Elsness 2009a: 92), and in non-continuative contexts (Meyer 1995: 224). It must be added that this supposedly characteristic AmE usage of the SPst can also be observed in other varieties of English and may be sensitive to effects of register and genre even within BrE (Miller 2004a: 244).

Another related point is that while some authors take it as self-evident that the relative frequency of the PrPf in contrast to the SPst has been decreasing in both BrE and AmE (e.g. Elsness 1997; 2009a; Yao & Collins 2012: 388), others state that the variation between the two forms is diachronically stable in these varieties (Hundt & Smith 2009). Yet others claim that the PrPf is expanding its territory at the cost of the SPst in some varieties, as it may be increasingly used in definite past time contexts (Engel 1998: 131). As noted above, a number of corpus studies of BrE and AmE have indeed identified instances of the PrPf co-occurring with definite temporal adverbials in both spoken and written material, which cannot all simply be attributed to performance errors or inadvertence on the part of the speaker (see further Section 6.6).

In particular, this section again tackles the issue of the PrPf/SPst-friendliness of different varieties, now in contexts involving indefinite/variable adverbials (*already, yet, always, ever, never, just, recently, since*). Further, in a similar fashion to the previous macro-level analyses, the influence of external factors such as variety type, geographical proximity and Americanization (in terms of lower PrPf-friendliness values) is tested.

Table 6.1. PrPf-friendliness scores across adverbials and varieties (assigned numerical values: + = 2, o = 1, - = 0)

	already	*yet*	*always*	*ever*	*never*	*recently*	*just*	*since*	PrPf-friendliness score
GB	+	+	o	o	o	o	+	+	1.33
IRL	+	+	-	-	-	-	+	+	1.00
AUS	+	o	o	o	o	-	o	+	1.13
NZ	+	+	o	o	o	-	+	+	1.25
CAN	+	+	o	o	-	-	o	+	1.13
PHI	-	-	+	o	o	-	-	+	0.75
JA	o	n/a	o	o	-	-	o	+	0.87
SIN	o	n/a	-	o	-	-	-	+	0.57
HK	+	n/a	o	-	-	o	+	+	1.11
IND	+	n/a	o	o	-	-	+	+	1.14
EA	+	+	+	o	o	o	+	+	1.50
NIG	+	n/a	o	o	-	-	o	+	1.00
PrPf-friendliness score	1.66	1.57	1.00	0.83	0.50	0.25	1.25	2.00	

Table 6.1 provides a brief overview of PrPf-friendliness values for each adverbial in each variety. These were calculated by establishing the average value for each line (variety) or column (adverbial) with a simple, underlying numerical scale that describes the relative frequencies of PrPf versus SPst co-occurrence for each item (+ means > 60% PrPf, assigned value: 2; o means PrPf = SPst, assigned value = 1; - means > 60% SPst, assigned value: 0). Average values may vary between 2 (= categorical PrPf co-occurrence/fully PrPf-friendly) and 0 (= categorical SPst co-occurrence/fully SPst-friendly). For the exact relative PrPf/SPst values of each item in each variety, see Appendix C and Werner (2013b: 217–228).

With the help of the PrPf-friendliness scores in Table 6.1, a hierarchy of adverbials with decreasing PrPf-friendliness along the lines *since* > *already* > *yet* > *just* > *always* > *ever* > *never* > *recently* can be established across all varieties: the further the items are to the right, the less

categorical the PrPf co-occurrence. Therefore, it can be concluded that all temporal adverbials that were considered more or less typical in PrPf contexts (apart from *since* constructions) in fact allow for variation to different extents or even correlate more strongly with the SPst. In addition, Wynne's reassessment regarding the typicality of adverbial PrPf co-occurrence in AmE and BrE that established *since*, *already* and *yet* as typical and *ever* and *recently* as less typical indicators of the PrPf (Wynne 2000: 111–112) can, in essence, be corroborated for the general picture across all varieties, which is also in line with Hundt et al. (2004: 567–568). Note further that adverbials that are semantically related (*yet*/*already*, *always*/*ever*) show similar average values, while adverbials of recentness (*just* and *recently*) diverge, and the negative element *never* does not align directly with the other items expressing indefinite frequency.

Second, from Figure 6.14 below, where the proportions between L1 and L2 varieties are compared with respect to the individual temporal adverbials, it might seem at first sight that the differences between the two variety groups are, globally, not overly pronounced. However, (i) the level of statistical significance for the difference between the L1 and L2 distributions is exceeded for all items apart from *recently*, *always* and *ever*,[289] (ii) with the exception of the last two adverbials, it is the L2 varieties that usually show a greater proportion of SPst co-occurrence, and (iii) some adverbials (notably *already*, *just* and *since*) correlate more positively with the PrPf than others (*never*, *recently* and *yet* in L2), while the distribution is nearly balanced for a third group (*always*, *ever* and *yet* in L1).

[289] Effect sizes vary. The exact values for the items with significant differences are as follows: *already* – $\chi^2 = 9.1139$, df = 1, $p < 0.005$, $\varphi_c = 0.088$; *yet* – $\chi^2 = 6.1188$, df = 1, $p < 0.05$, $\varphi_c = 0.233$; *never* – $\chi^2 = 4.0513$, df = 1, $p < 0.05$, $\varphi_c = 0.050$; *just* – $\chi^2 = 33.0922$, df = 1, $p < 0.001$, $\varphi_c = 0.190$; *since* – $\chi^2 = 7.9421$, df = 1, $p < 0.005$, $\varphi_c = 0.132$.

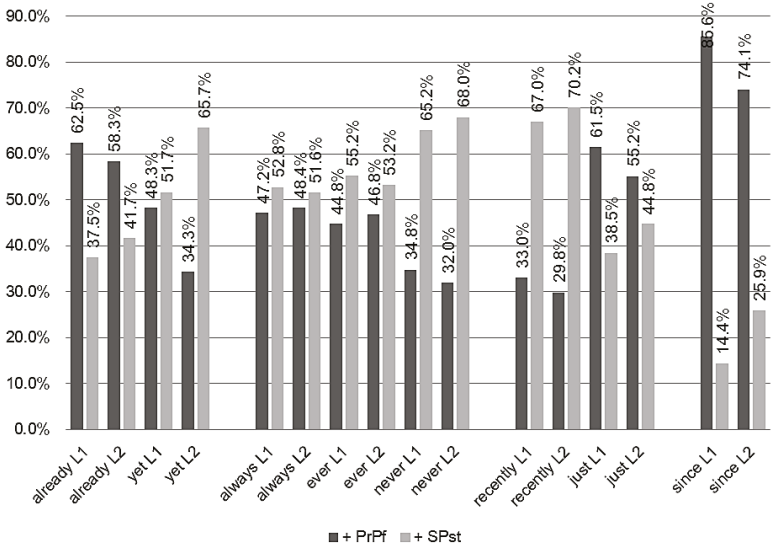


Figure 6.14. Relative distribution between co-occurring PrPf and SPst forms for each adverbial: L1 vs. L2 varieties

Table 6.1 above also enables us to establish the following hierarchy of PrPf-friendliness for the varieties: EA > GB > NZ > IND > AUS = CAN > HK > IRL = NIG > JA > PHI > SIN. BrE and EAfE emerge as the varieties with the highest proportions of PrPf co-occurrence, followed by the Antipodean varieties and CanE. The alignment of EAfE and, moreover, IndE (and potentially also HKE) with the more PrPf-friendly L1 varieties may be interpreted as evidence of continuing exonormative orientation in these varieties. The transplanted L1s cover the middle ground, while PhiE, JamE, SinE and, less distinctly, NigE can be categorized as varieties that reveal a clear overall SPst preference.[290] It is noteworthy that these are the varieties that had and have close relations with the US, either politically (PhiE), geographically (JamE) or economically (PhiE,

[290] Hundt & Biewer (2007: 257–259) provide other measures of PrPf-friendliness. One is based on the Mossé-coefficient that revealed a split between more friendly L1 varieties (AmE, BrE, AusE, NZE) and less friendly L2 varieties (PhiE, SinE and Fiji English). The other (relative frequencies of the PrPf as potentially equivalent to the SPst) was inconclusive, as extrapolation of manual counts led to skewed results.

SinE, JamE). Learner effects, such as the generalization of the structurally simpler SPst, seem to play a secondary role only in this instance, as shown by the alignment of other L2 varieties (EAfE, HKE, IndE) with L1 varieties.

The findings also closely relate to Schneider's (2007) phase model of postcolonial varieties (see Section 2.5.1 and the discussion in Section 7.3 and Werner in preparation a). The L2 varieties on the more PrPf-friendly pole (EAfE, HKE and IndE) are in phase 3 (nativization), while the SPst-friendly ones (PhiE, SinE, and JamE) are in transition to or have already entered phase 4 (endonormative stabilization). In addition, the non-emergence of the Antipodean varieties as new linguistic epicenters for Asian and Pacific varieties can be corroborated beyond the evidence provided by Hundt & Biewer (2007).

Finally, to test whether the hierarchy of PrPf-friendliness for the varieties that is based on the above measure is meaningful, hierarchical cluster analysis is employed, taking into account all quantitative data points.[291] This helps us to establish a measure of similarity between the varieties and to identify clusters of varieties that behave similarly. Figure 6.15 below presents the tree-shaped dendrogram (c = 0.61) created with R.

Two clusters emerge from the data. Cluster A comprises SinE, PhiE, JamE and NigE; that is, those varieties that show higher proportions of SPst co-occurrence and that have reached phase 4 in Schneider's (2007) model (apart from NigE). This might be labeled the "SPst prevalence" cluster. The remaining varieties constitute cluster B, the "PrPf prevalence" cluster, which can be subdivided into two subclusters. Cluster B1 comprises BrE and the L2 varieties that are in phase 3 of Schneider's (2007) model and arguably still show exonormative orientation toward BrE, while B2 contains the transplanted L1 varieties. In sum, the findings from the cluster analysis indicate that variety type and extralinguistic factors also influence the distribution between the PrPf and the SPst for the group of adverbials under investigation. A comparison with the variety hierarchy established above reveals some slight deviations, for

[291] Note that *yet* was excluded due to the low token counts for this item, which may result in a disproportionately strong effect on the outcome of the cluster analysis.

example as regards the closeness of the Antipodean varieties to one another and as regards the immediate clustering of the L2 varieties in phase 3 with BrE. On the whole, however, the cluster analysis confirms the above picture.

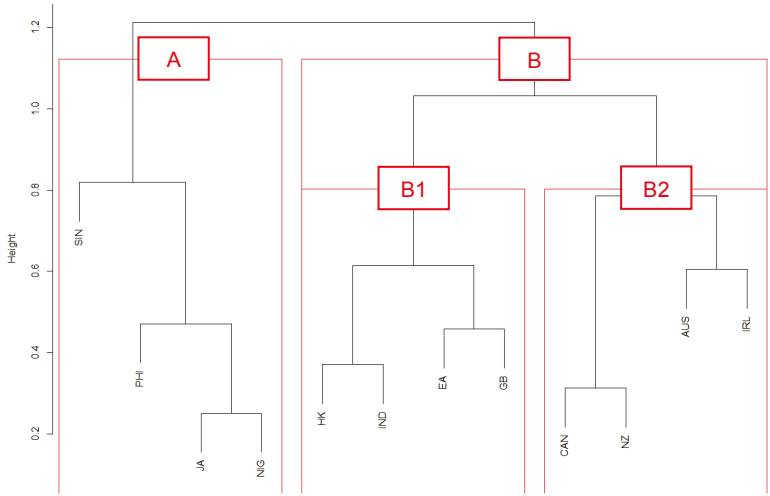


Figure 6.15. Cluster dendrogram of the overall similarity of the ICE components: temporal adverbials

The findings of the cluster analysis are verified by the network representation as shown in Figure 6.16. Again, the SPst-friendly/phase 4 varieties (SinE, PhiE, JamE and NigE) are close to one another and contrast with the transplanted L1/phase 5 varieties (CanE, NZE, AusE and IrE) as well as with the PrPf-friendly varieties (BrE + phase 3 varieties HKE, IndE and EAfE). Note that BrE is also relatively close to the phase 5 varieties in this non-hierarchical representation, and that the distance between the two Antipodean varieties is smaller than could be assumed from the hierarchical cluster analysis, where AusE and IrE appear similarly distant from NZE.

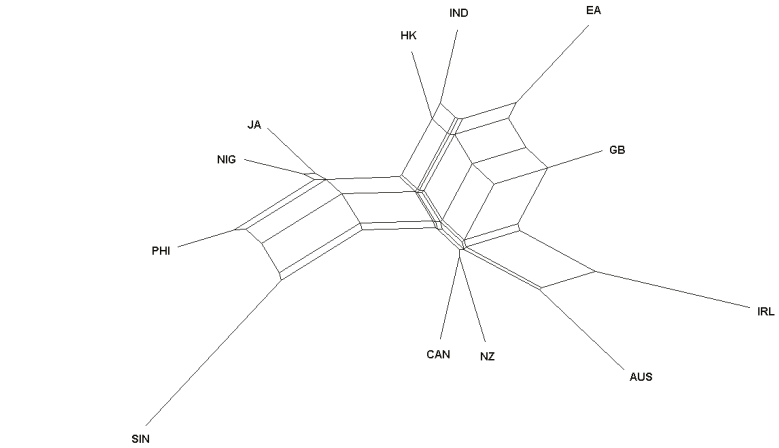


Figure 6.16. NeighborNet of the overall similarity of the ICE components: temporal adverbials

6.6 Case study 3: The Present Perfect as a narrative tense

This section explores whether some ICE varieties show a tendency to use the PrPf as a tense (*sensu* Quirk et al. 1985),[292] again in the environment of temporal adverbials. While this phenomenon in general is salient in a number of varieties of English (Platt et al. 1984: 71; eWAVE Feature 100 2012), with the "personal or emotional involvement" of the speaker said to be a potential trigger (Walker 2011: 78), it has not been tested whether it also appears in educated English as represented in ICE. Accordingly, a corpus search for PrPf instances that violate the definite temporal adverbial constraint was conducted;[293] that is, occurrences that are considered ungrammatical or odd in standard BrE were identified in the ICE data (see Section 3.3 above for theoretical issues). These items, which usually co-occur with the SPst, could be categorized as typical "+ then" adverbi-

[292] For a more detailed discussion of terminological issues and the suitability of standard reference grammars for worldwide varieties of English in the area of the PrPf, see Werner (2013a).

[293] The search was for a selection of adverbials (x + *ago*, *once*, *yesterday*, *last* + x, *in* + cardinal number) that indicate a specific time in the past co-occurring with the PrPf.

als in line with McCoard (1978: 135) or, alternatively, as "– current relevance" in Davydova's revised scheme (2011: 69). Note that McCoard's/Davydova's lists also consider the temporal adverbials *in the past* to be atypical with the PrPf. However, the present search of the ICE components revealed that *in the past* freely co-occurs with the PrPf across all varieties and genres, for instance as in (1) or (2).

(1) there have been some curling associations that have complained about the provincial identification that one *has associated in the past* with the colour (ICE-CAN s2a-018)
(2) This move does not in any way affect the business that you *have done in the past* with the MI Group. (ICE-GB w1b-016)

Therefore, *in the past* was excluded from the current search and I propose that it should be seen as an adverbial that merely expresses indefinite past, which might or might not have repercussions for the present moment/current relevance. In turn, this suggests that it should be categorized in the "± then/current relevance" group.[294] This matter apart, the current analysis allows us to test the hypotheses that (i) L2 varieties are more flexible in their usage of the PrPf versus the SPst in these contexts, too, and (ii) the PrPf is used as a tense that is on a structural/functional par with the SPst.

Table 6.2 presents the absolute figures for definite past adverbials co-occurring with the PrPf in the dataset.

[294] Wynne's data indeed suggest *in the past* as a typical PrPf trigger (2000: 111–112). Further contexts where the constraint is suspended do occur (see also examples (2) to (5) in the preceding section), e.g. with afterthoughts/insertions such as *They have also uhm I think last year uh given a list of of six principles* (ICE-IRL s2b-001), in iterative contexts such as *For example, the diagnostic and statistics manual (DSM) has been updated twice, once in 1968, and again in 1980, with a revised version appearing in 1986* (ICE-GB w1a-007; see further Engel 1998: 132) or with postponed temporal specification due to pragmatic necessity as in *I've written a letter* (speaker B) *Have you* (speaker A) *Yeah last week* (speaker B) (ICE-IRL s1a-059). Pauses and hesitations may also play a part (Bowie et al. 2013; cf. example (20) below).

Table 6.2. Co-occurrence of PrPf and definite temporal adverbials in ICE

	spoken	written
GB	1	2
IRL	3	1
AUS	1	0
NZ	4	0
CAN	1	0
sum L1	*10*	*3*
PHI	5	2
JA	2	0
SIN	3	1
HK	11	4
IND	10	2
EA	2	1
NIG[295]	1	0
sum L2	*34*	*10*
total	*44*	*13*

The first noteworthy finding is that combinations of the PrPf and definite adverbials are attested across all varieties, although absolute numbers are low, which suggests that it constitutes a rare phenomenon in the majority of the L1 and L2 varieties contained in ICE. Despite the low absolute numbers, some tentative conclusions can be drawn. First, PrPf co-occurrence is apparently more common in spoken data. This could be expected, as the spoken mode has been found to be more susceptible to this innovative pattern[296] elsewhere (Hundt & Smith 2009: 55–56). Second, overall there is no significant difference between the L1 and the L2 variety group.[297] However, within the L2 group, the Asian varieties IndE and HKE and, to a lesser extent, PhiE are conspicuous with regard to their relatively high absolute frequency of PrPf co-occurrence with defi-

[295] Speech is reported here (ICE-NIG reportage Pr_11); therefore, the example is categorized as spoken.

[296] As co-occurrence of the PrPf with definite temporal adverbials is salient in older stages of English (Miller 2004a: 235), it is in principle also conceivable that it represents a remnant that was introduced in early stages of colonization and has been preserved in postcolonial varieties.

[297] Due to the low number of occurrences in some of the cells, Fisher's exact test was employed for significance testing.

nite adverbials, while the remaining L2 varieties align with the L1 varieties. These results tie in with Lunkenheimer's (2012: 338, 340) finding that PrPf usage in SPst contexts typically occurs in Asian varieties, which are characteristically of the L2 type.

Different factors may serve to explain the outliers. The PrPf has traditionally been prevalent in all contexts in IndE, and it has been suggested that processes of transfer (re-analysis of the Hindi structure past participle + auxiliary and extension of its application to PrPf contexts) and substrate influence (for speakers of Dravidian languages) play a role for this variety (Davydova 2011: 172–173, 190), resulting in examples such as (3) and (4).

(3) He's been good *yesterday* (ICE-IND s1b-047)
(4) You know *yesterday* I have seen some two guys speaking with you (ICE-IND s1a-049)

In the absence of contradictory evidence and as no speaker information is available for the ICE components, this hypothesis seems to provide a plausible explanation. A similar line of argumentation is pursued by Sharma (2001: 367), who suggests that the PrPf in IndE may be developing from PERFECT to PERFECTIVE; that is, expressing completion of an event rather than continuing relevance. If this is correct, the PrPf in IndE would follow a common typological path of development (see Section 2.1 above).

For HKE, it has been noted that its speakers commonly do not mark verbs for tense or aspect (or only draw a distinction between present and past tense), creating time reference through temporal adverbials alone (Platt 1982: 410; Setter et al. 2010: 49–55; cf. Section 5.9). Therefore, it is not surprising that the corpus data contain examples of this kind, such as (5) or (6).[298]

(5) And then uhm but *recently* her backbone is very painful [...] (ICE-HK s1a-052)
(6) [...] we *just* have a quarrel *last night* (ICE-HK s1a-035)

[298] In these two examples, indefinite adverbials are used.

It could be argued that perfect marking through a combination of a TR-form and a temporal adverbial is indeed redundant and thus is avoided. From the perspective of the speakers, who use a pattern familiar from their L1, this does not result in any loss in communicative value.

Greater variation in grammatical forms, including the PrPf with definite temporal adverbials as in (7) can also be observed (see further below).

(7) Some of them have *once* been my best friends [...] (ICE-HK w1b-004)

In addition, forms that are partly marked (*I've mention*) as in (8) occur.

(8) So the most important thing when you look at the prologue okay is what as I've mention *last time* is the fact that it gives you a time structure [...] (ICE-HK s1b-010)

Here, the same argument as above is valid. The usage of this reductive pattern does not result in any loss of communicative value. It rather is a pattern approximating standard usage and occurring regularly in the data (see also Section 6.4), now in combination with a definite temporal adverbial.

All the foregoing items are again most likely due to a combination of substrate influence (Kachru & Nelson 2006: 42) and learner effects – a common scenario in L2 Englishes (Schneider 2012: 63–64; 85; cf. the contributions in Mukherjee & Hundt 2011) – and surface also in the ICE-PHI data, as in (9).

(9) *Months ago* I have written Sen John Sheffield the head of the US Senate Environment Committee about the problems [...] (ICE-PHI s2b-032)

Looking at individual examples, a qualitative difference between the occurrences in the L1 varieties and the L2 varieties can be established. In both variety groups the subjective conception of a situation as recent and

relevant by the individual speaker may be conveyed by a combination of definite and indefinite temporal adverbial as in (10) (see (11) for a corresponding SPst example) or by a premodification (*this last week*; *only yesterday*) as in (12) or (13).

(10) I would not say that women's issues have *just* started *last year* I would not even say that (ICE-EA Kenya s1a-028)
(11) I *just* had this kind of patient *a couple of months ago* (ICE-PHI s2a-032)
(12) Oh I've had some fun *this last week* (ICE-CAN s1a-093)
(13) Indeed my English is likely to be a little dodgy as I have *only yesterday* returned to civilisation after a 3-week sojourn to Derry (ICE-IRL w1b-006)

In contrast, the vast majority of occurrences in the L2 varieties clearly represent definite contexts that usually require the SPst, as in (14) to (18).

(14) Actually I've noticed that *two weeks ago* (ICE-PHI s1a-066)
(15) This work came out of a discussion that Noah and I have had *several months ago* probably in October uh last year when we were talking about ethics [...] (ICE-JA s2b-029)
(16) So he admits Sir what he has stated *yesterday* was was not correct (ICE-IND s2a-063)
(17) Where have you been *last Sunday* (ICE-HK s1a-085)
(18) We have mailed you the above DBS Card *2 weeks ago*. (ICE-SIN w1b-019)

Thus, examples such as (14) to (18) may well be explained by substrate and learner influence that gives rise to an innovative pattern; that is, the possibility of replacing the SPst with a PrPf form in indefinite contexts. This arguably leads to a conception of the PrPf as a tense that is on a functional par with the SPst in these varieties. It needs to be borne in mind, however, that these occurrences are also quantitatively rare overall in the L2 varieties.

For the L1 varieties, the few examples where the PrPf co-occurs with definite temporal adverbials can be explained in terms of pragmatic necessity, as argued for example by Rastall (1999: 81–83; cf. Binnick 2009: 272) and Hundt & Smith (2009: 58), almost without exception (but see (19) and (20) from AusE and NZE, two varieties where PrPf use in typical SPst contexts is rated "neither pervasive nor extremely rare" in eWAVE Feature 100 2012).[299] Therefore, it seems unlikely that a general development toward PrPf-SPst convergence/interchangeability will ensue in the L1 varieties in the foreseeable future.

(19) Well he's come on very quickly *last year* (ICE-AUS s2b-017)
(20) well because i mean we can read up on everything else again but we've seen that video *months ago* (ICE-NZ s1b-009)

Even if PrPf usage in combination with definite temporal adverbials becomes accepted in one variety (e.g. one of the L2 varieties or one of the transplanted L1 varieties such as AusE; see Ritz 2010), whether this development will gain currency over time and eventually spread to other varieties remains a matter of speculation.

[299] Note that in example (20) the corpus annotation contains a hesitation marker between *months* and *ago*. This may be interpreted as an instance of the speaker becoming aware of a "performance error", but accepting the lack of grammatical well-formedness for pragmatic reasons.

7 Discussion of the findings

The central purpose of this section is to discuss and condense the main points emerging from the results of the empirical study above (chapters 5 and 6). At the same time, the discussion section reconsiders the issues raised at the outset of the study (chapters 2 to 4) in light of the data presented in the preceding chapter. Thus, the corpus findings are further related to the research context as well as to broader theoretical issues.

7.1 Overlap or divergence?

This section addresses the first guiding research questions of the study in that it considers overlap and divergence of worldwide varieties of English and discusses the issue of the PrPf as a core feature of English grammar. Its main aim is to further contextualize the empirical findings of sections 6.1 to 6.4.

First, the analysis suggests that, in general, the PrPf is a globalized or core structural feature that is relatively uniform across varieties of all types (see sections 6.1 and 6.2), despite some quantitative differences in usage patterns (see sections 5.1 to 5.10). On a related note, the overall result that has emerged from the present analysis, namely that differences between "old" L1s, "transplanted" L1s, and L2s are quantitative rather than categorical in nature, is in line with the outcomes of previous corpus-based studies of morphosyntactic phenomena in World Englishes (cf. e.g. the contributions in Rohdenburg & Schlüter 2009, Mukherjee & Hundt 2011, Hickey 2012 or Schreier & Hundt 2013). This supports Hundt & Vogel's suggestion of an alternative means of categorization, namely into "set[s] of varieties that show (a) the effect of language contact and (b) exo-normative influence to a greater or lesser extent" (2011: 161; see further Section 7.3), also given that in the present dataset varieties with continuing exonormative orientation toward BrE, such as HKE or IndE, occupy leaves and nodes adjacent to their "parent" variety, while others, such as PhiE, which already have entered the phase of endonormative stabilization to some extent, cluster more closely with

the transplanted L1 varieties. Note that the role of Americanization needs further consideration in this respect.

Furthermore, patterns of usage appear to be sensitive to the external variable "text type"/"register". The dendrogram and NeighborNet analyses showed that the vast majority of spoken texts and many spoken subcategories, for instance private dialogues (s1a), were to be found in adjacent leaves or nodes and were clearly not determined by variety type, a finding that corroborates the result from the preceding multidimensional analyses. For written texts the picture was less clear, with the members of some categories in the same clusters, while others were distributed more widely. Overall, this suggests that text type may represent an external factor that determines the patterns of usage of the standard PrPf irrespective of the variety or variety type. This finding further supports the claim that similarities between text types are stronger than between varieties (e.g. AmE and BrE; Schlüter 2002a: 109) and provides counterevidence to assertions that the similarity of text types across variety types is a "myth" (Kachru & Nelson 2006: 283). Secondly, the aggregational representations revealed considerable variation within individual text categories, as illustrated by the outliers and split clusters of the same text category. By and large, these findings corroborate earlier corpus-based register studies, which found that, in general, written texts possess greater potential for variation (in our case, for usage patterns), while spoken texts are more limited and, thus, prone to sharing similar characteristics (Biber & Conrad 2009: 261–262).[300] This also underscores the validity of pre-established textual categories as concepts that are based merely on cultural conventions rather than language-internal properties.[301] All in all, however, the pre-established cate-

[300] Biber & Conrad provide a number of "highly informational written registers produced for experts in a field – registers like academic research articles or official documents" (2009: 261) that share distinctive linguistic features. As regards PrPf usage, such distinctive written registers that align closely may be represented in the current data by academic writing (w2a) or creative writing (w2f).

[301] It is taken as self-evident that establishing textual categories based solely on language-internal features is a tedious task that necessarily involves subjective decisions, for instance with regard to the inclusion of the structural features that are part of the analysis. A relevant research project attempting a re-classification of the texts of the British National Corpus is currently underway (Ghada Mohamed, University of Lancaster, p.c.).

gories seem to mirror actual inherent structural differences for the PrPf usage in some ways, at least as regards the spoken versus written distinction and a number of categories.

A note of clarification is in order here. Some of the contextual variables that were coded could be seen as inherent general properties of one textual category or another and not as related to PrPf usage in particular. For instance, a high(er) proportion of questions and preceding PrTs forms may be interpreted as typical of dialogues in general as an interactive and present-oriented category. Thus, it is arguable whether associations between (particularly spoken) text types, as depicted in the multidimensional analyses (both in the individual varieties and in the overall view), are due to general text type effects rather than genuine similarity in PrPf usage patterns. However, associations between spoken text types were shown to persist even when the allegedly general properties (sentence type and/or preceding TR form) are disregarded in the multidimensional analysis.[302]

Despite the register and genre influences identified, more often than not, significant differences had small effect sizes or could be found on lower levels or in tendency in the groupings only. The multidimensional analyses yielded no clear-cut divisions along variety types, modes of discourse, macro-genres or ICE text types. The only language-external dimension of influence could be found in the form of an "involved" versus "informational" split (see Section 6.3), which again indicates the applicability of these categories. But a post-hoc test revealed no clear-cut result, which provides further support for the hypothesis that, when individual grammatical features are considered (here, the PrPf), a neat categorization according to variety type is impossible or at least inadequate, given the blurred distinctions between L1, transplanted L1, L2 and arguably also learner varieties of English, potentially due to the fact that all contact scenarios are "messy" (Schreier 2010: 464) and involve

[302] This is exemplified with a hierarchical representation for one variety (HKE) in Appendix D, where the values for each individual variable are disregarded step by step. Even if sentence type *and* preceding TR form are left out (see Figure 8.f.), the spoken categories are located on neighboring leaves (with a significant monologues subcluster). A similar result occurs when taking into account the leaves for all varieties (as in Section 6.2).

processes of linguistic globalization (Hundt & Vogel 2011: 161).[303] These results also point to a restricted influence of further language-external variables and, in consequence, a categorization of the PrPf and its grammar of usage as a core structural feature of worldwide varieties of (educated) English as represented in ICE.

That only minor divergence was found between varieties and variety types does not come as a major surprise. First, from a sociolinguistic point of view: despite its political connotations, Fishman's coinage "anglification" (1996b: 637; see further Section 2.5.2 above), which implies a continuing exonormative orientation and thus relative linguistic homogeneity across speakers, appears to be a helpful concept. Anglification, and thus homogeneity, is most advanced at the higher end of the social and educational scale and, due to the corpus compilation principles, this is where the majority of speakers who contributed to the ICE data can be located. As a result, divergence should become more pronounced if we employ data from mesolectal or basilectal – and therefore probably less anglified – speakers of English. However, the analysis has also shown that "educated" by no means equates to "standard", let alone exonormative orientation toward BrE in all respects, as was indicated by the usage of alternative surface forms and potential extension of the PrPf territory. Additionally, effects of register and genre were also shown to play a part. Second, from the point of view of linguistic subfields: While it has been observed that on the level of phonology spoken varieties of English are susceptible to increasing divergence and fragmentation (see e.g. Wächtler 1977: 64; Leitner 1992: 219), grammar (in terms of morphology and syntax) has rather been seen as a monolithic block that is relatively immune to contact influence (Sankoff 2001: 658) or, alternatively, as a linguistic area where patterns increasingly converge internationally (Hundt et al. 2004: 588). Still, subtle structural differences between varieties, variety types, etc., do exist. In the present study these mainly consist of quantitative differences in the relative distributions of

[303] Görlach mentions a related aspect, stating that the historical BrE or AmE "ancestries" of varieties of English "are becoming diluted and probably will ultimately be lost" (2002: 117). See further Omoniyi & Saxena (2010: 5) or Ramanathan et al., who argue for a radical reconfiguration of the traditional categorizations of varieties of English as globalization "render[s] much of our discipline's polarities around English [...] defunct" (2010: xvii).

single language-internal variables, but also of qualitative differences in terms of alternative forms (see Section 6.4 and below). In the multidimensional analyses, these differences emerged in the form of distances between varieties, variety types, etc., which were often relatively similar, such that we may speak of a "similarity in differences". In other words, while all varieties are relatively uniform overall, differences could still be discerned when looking at their "inner life". This is mainly due to the fact that the analysis of aggregated data may naturally result in some loss of detail (but cf. sections 5.1. to 5.10 above). It has to be kept in mind, however, that even if PrPf usage is characteristic of particular registers and text types, it is not a sufficient condition for distinguishing between them, as patterns overlap across language-external categorizations on all levels.

In contrast to the aspects mentioned so far, the case study in Section 6.4 deliberately did not consider regularizing usage but instead investigated "exotic", that is, alternative, surface forms in perfect contexts in order to determine similarities and differences across varieties. As regards specialized semantic or pragmatic values of these forms, the results remain inconclusive at first sight, with none of them seemingly restricted to specific contexts. However, a possible exception is the *after*-perfect in IrE, which expresses recentness, while no evidence could be found for any other alleged pragmatic function such as "mirativity" or "solidarity among speakers". Still, it has to be acknowledged that this form (like the other non-standard forms) is rare overall in ICE.[304] Different sets of (spoken informal) IrE data would allow for a more comprehensive qualitative analysis and might further contribute to solving the issue of the tenability of the (semantic) grammaticalization hypothesis of IrE perfect forms (Kallen 1989 vs. Pietsch 2009; cf. above). While it is in principle conceivable that particular semantic functions could be assigned to different surface variants, as has indeed been observed in varieties such as Trinidadian English (Winford 1993: 141), the ICE-IRL data provided no evidence for the grammaticalization hypothesis. A closer investigation would probably necessitate the involvement of other types

[304] This equally applies to other corpora, for instance the data used by Fritz (2006).

of data, e.g. questionnaires or sociolinguistic interviews (see Krug & Sell 2013 for an overview of questionnaire-based research).

For the remaining alternative surface forms,[305] the present findings suggest that these do not carry idiosyncratic semantic notions but rather exemplify vernacular features across a number of varieties, and may therefore be analyzed as "angloversals" (Szmrecsanyi & Kortmann 2009: 33)[306] or, alternatively, as some kind of "genre/register versal" within English. This claim is substantiated by the facts that (i) relevant forms occur rarely in ICE (see also Seoane & Suárez-Gómez 2013), which represents educated usage, and (ii) they occur almost exclusively in the spoken sections of the corpus data. Therefore, the different realizations are due to register effects rather than semantic or pragmatic influences in the narrow sense, which is why they can be used across a variety of contexts. However, a corresponding, but rather general pragmatic value that could be identified for the majority of the relevant forms is that of "conveying informality".

7.2 Present Perfect versus Simple Past

This section focuses on the wider implications of case studies 2 and 3; that is, the combinability of the PrPf and the SPst with temporal adver-

[305] See Section 6.4. The origin of these forms was not discussed in detail above. Suffice it to say that, from a diachronic perspective, two sources for non-standard surface realizations are conceivable. They may be due to substrate influence (as claimed for IrE) and learner effects/cognitive constraints (applies mainly to L2 varieties; cf. e.g. Seoane & Suárez-Gómez 2013: 11) or, alternatively, persistent influence from earlier and non-standard varieties of English (e.g. traditional dialects), where these forms have been salient. Naturally, these two dimensions interact, such that earlier learner effects and/or substrate influence might eventually be transferred to L2 varieties and interact there with other learner and substrate effects (see also Hickey 2010a and the contributions in Schreier & Hundt 2013). To determine the exact nature of these interactions empirically, we would need both synchronic and historical data of all varieties involved in the process. While these are more widely available for earlier stages and dialects of the English spoken in Britain, projects on the establishment of historical corpora for L2 varieties are in a fledgling state only (but cf. "The Corpus of Historical Singapore English" project; Hoffmann et al. 2012).

[306] But cf. Winford (2009: 233; 2013: 227), who rejects the claim that tense-aspect systems of non-native varieties of English in their entirety represent some kind of typological universal.

bials (see Section 6.5) and the use of the PrPf as a narrative tense (see Section 6.6).

Above all, the analysis of temporal adverbials and their co-occurrence patterns with either the PrPf or the SPst across the extended set of twelve varieties of English (i) has helped to shed more light on the issue of variation between the two grammatical forms, (ii) complements previous studies that were entirely or primarily based on AmE and BrE, and (iii) addresses the issues discussed below with the help of a quantitative analysis of ICE data.

First, as regards the patterns of variation between co-occurring PrPf and SPst with a number of temporal adverbials, it emerged that the indefinite temporal adverbials under investigation by no means represent a homogeneous group. A hierarchy of PrPf-friendliness was established for each item, which revealed that PrPf co-occurrence is almost categorical with some adverbials (e.g. *already*, *yet* and *since*), while others are more flexible (e.g. *always*, *ever* and *just*) or even co-occur largely with the SPst (e.g. *never* and *recently*). Globally, however, a wide range of distributional patterns was observed across varieties as well as within each individual variety for the items considered. Findings from earlier studies that were based on AmE and BrE data only or that covered fewer varieties or a small selection of texts from a few ICE components were extended and qualified (see also Werner 2013b). On the whole, Hundt & Smith's (2009: 58) assertion that variation between the PrPf and the SPst with indefinite temporal adverbials represents a case of layering can be confirmed for further varieties, although the alleged stability of this variation would have to be confirmed by further diachronic studies, especially for transplanted L1 varieties and L2 varieties.[307]

Second, as regards individual varieties, perceptions such as BrE being a PrPf-friendly variety in temporally specified contexts could be confirmed, while L2 varieties could be found near two poles of a hierarchy of PrPf-friendliness (see also sections 5.1 to 5.10 and 6.4). This sug-

[307] See also Kortmann & Szmrecsanyi (2004: 1188–1189), who found leveling between the PrPf and the SPst among the common features of L2 Englishes. As leveling between the forms also occurs in a range of other non-standard varieties of English, Schneider (2007: 84) proposes that this feature qualifies as a vernacular universal.

gests that some (EAfE, IndE and HKE) are still exonormatively oriented toward BrE, while others (PhiE, JamE and SinE) show a more SPst-friendly, arguably "Americanized", pattern[308] that can be attributed to close geographical (areoversal) or intensive political, cultural and economic relations. In light of the fact that other L1 varieties are also less PrPf-friendly than BrE, it may be argued that SPst-friendliness or potential for variation with temporal adverbials has become a globalized rather than an Americanized pattern. However, variety types (*sensu* Schneider 2007) do also exert some influence here. This claim is substantiated by the findings from the aggregational quantitative analysis where BrE and the exonormatively oriented L2 varieties (phase 3), the transplanted L1 varieties (phase 5) and the SPst-friendly varieties (phase 4) cluster together. The non-hierarchical NeighborNet further showed that the distances between BrE and the transplanted L1s are shorter than the distances between L1s and the L2 varieties in general. Therefore, we can establish a split between the L1 and L2 varieties, while differences between the traditional L1, BrE, and the transplanted L1s are less marked.

As regards the weakening of the definite temporal adverbial constraint and the usage of the PrPf as a narrative tense, this indeed occurs in definite temporal contexts in the ICE data, though rarely overall and predominantly in spoken genres or involved texts such as letters (w1b). With respect to the objection that the low number of occurrences may be merely ascribed to the compilation principles of ICE, which captures mainly non-vernacular English, note that Davydova's (2011) data on mesolectal and basilectal varieties, which were obtained with the help of sociolinguistic interviews, also did not yield sufficient numbers to attempt a meaningful quantitative analysis. A qualitative analysis revealed that while pragmatic pressure on the speaker as proposed by Rastall (1999) may override the constraint in L1 varieties, L2 speakers are more likely to show creative usage of the PrPf, that is, to use it interchangeably with the SPst, due to substrate or learner effects or a combination of

[308] It has already been indicated that, as yet, no complete ICE component that would allow for direct comparison is available for AmE, but previous studies have shown that, overall, AmE is much less PrPf-friendly than BrE, although distributional values vary considerably for individual items (Hundt & Smith 2009: 53–54; see also Yao & Collins 2012).

both. Therefore, we can conclude that in this respect, too, there seems to be a clear split between L1 (both traditional and transplanted) and L2 varieties. Based on the present evidence it is, in principle, conceivable that change in this grammatical area (development of the PrPf into a tense on a structural par with the SPst, i.e. a fully-fledged variant) might occur (cf. the typological evidence in many other languages; Miller 2000: 350). However, at present the case for an extended functional range of the PrPf in the educated varieties under investigation is weak and no definite answer can be given as to whether this will materialize.

7.3 Models of World English revisited

As indicated above (Section 6.1), the traditional conceptualizations of varieties of English leave room for improvement. In the present section I thus elaborate further on the topic of models of World Englishes. In particular, I seek to assess the applicability to grammatical studies of some of the models presented in Section 2.5.1, namely Kachru's (1988) approach as the most widespread circle model, as well as the models developed by Modiano (1999), Minnich (2004) and Schneider (2007). I also discuss some more general implications of models of language contact.

In relation to Kachru's approach, the findings on the PrPf from the ICE data support the position that the ENL/L1 versus ESL/L2 (versus EFL) distinction is of limited explanatory power when it comes to the investigation of individual features (as argued e.g. in Hundt 2009). In the quantitative analyses involving aggregated data (hierarchical dendrograms and non-hierarchical NeighborNets), clear splits along the lines of variety types were not in evidence (see sections 6.1 to 6.3 and 7.1) or could be observed merely in tendency. A similar result emerges even when applying the more fine-grained categorization proposed by Kortmann & Szmrecsanyi (2011), which distinguishes between standard varieties (BrE/AmE), transplanted L1s, shift varieties and non-native L2s. To take the above position even further, it seems that Schneider's (2007) influential dynamic phase model is applicable when considering diachronic and sociocultural aspects (as explained in the background sec-

tions on each variety; see Chapter 5) such as identity issues and related aspects rather than when investigating structural features in isolation. In this regard one of the truisms of variational study becomes apparent here, namely that people and not nations use languages (Conrad 1996: 22). This also highlights the discrepancy between socially and politically motivated and inherently linguistic criteria when labeling certain varieties (see further Spolsky 2009: 1 on this topic).

However, there is also evidence that the L1 versus L2 distinction and Schneider's (2007) model should not be dismissed *in toto*. This was revealed in the assessment of the PrPf as a narrative tense, where clear qualitative differences between L1 and L2 varieties emerged in combinations of the PrPf with definite temporal adverbials (see sections 6.6 and further 7.1) and, likewise, during the quantitative analysis of the PrPf-friendliness of various temporal adverbials (see sections 6.5 and 7.2). In the latter case study, PrPf-friendly varieties were shown to comprise BrE and L2s that have advanced to phase 3 only, while transplanted L1s (phase 5) covered the middle ground (but were still close to BrE) and L2s that have already reached phase 4 were least PrPf-friendly, arguably showing traces of Americanization (or, globalization, depending on the researcher's perspective). Thus, both Kachru's circles and Schneider's phases are useful as they also involve and are able to mirror actual structural conditions, albeit with a limited scope.

With regard to the applicability of general processes as identified in language contact studies (e.g. Thomason 2001; 2010 or Muysken 2010), the relative homogeneity of the varieties in the area of the PrPf runs counter to some of the expectations formulated above (see Section 2.5). Contact between typologically distant languages (as in the L2 varieties) did not result in overall patterns that made these varieties markedly different from L1s. However, as indicated in the foregoing discussion of Kachru's and Schneider's models, a number of qualitative and quantitative differences between the two types and, moreover, between individual varieties do exist. To mention a relevant example, with the use of *already* or *ever* as sole perfect markers (see e.g. sections 5.8 and 5.9), the present data illustrate that "under circumstances of intense contact, any linguistic feature can be transferred to any other language" (Thomason 2010: 41), which is usually uncommon to happen for grammatical fea-

tures. This indicates that these general processes of language contact should not be overlooked in studies of individual grammatical features, even if their explanatory potential is limited in the present study with regard to the potential effects of the geographical closeness of varieties.

Due to the above limitations, I suggest that for the description of individual structural features in model form, Modiano's (1999) approach seems to represent the most promising candidate. Its graphical representation is shown once more in Figure 7.3.1 for convenience.

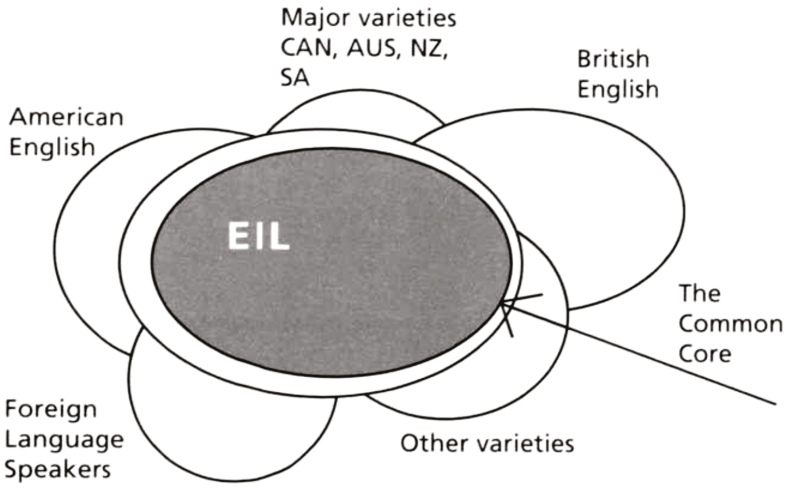


Figure 7.3.1. Modiano's (1999) model of World Englishes

While "English as an international language" (EIL) is admittedly only a vaguely defined concept, lacking universal acceptance in the World Englishes community due to its being more of a theoretical construct than an actual variety, the main advantage of Modiano's model compared to Kachru's and Schneider's models is its flexibility. Further, it is more comprehensive and concrete than alternative suggestions as proposed by Hundt & Vogel (2011: 161; "set[s] of varieties that show (a) the effect of language contact and (b) exo-normative influence to a greater or lesser extent") or the "Variety Spectrum" approach as sketched in Bongartz & Buschfeld (2011: 45–48), where merely the variation of single features

and its interaction with sociolinguistic categories can be plotted in a two-dimensional way. The flexibility of Modiano's approach is grounded in the fact that individual features can be placed not only within the darker core (EIL) area, but also in one or even all of the overlapping circles, irrespective of the variety status and types that these circles are associated with. In addition, the approach allows for idiosyncratic items to be included in one (or some) of the outer circles, which again are potentially overlapping. While it is clear from the analyses above that the PrPf would be located in the core, alternative surface forms (such as those described in Section 6.4) could be placed in one (or more) of the outer circles, to which a circle labeled "indigenized" L2 varieties should be added. Conceivably, the relationship between the variety types could be displayed more equally, resulting in a flower-like graphical representation. Further petals could then also be added, such as "other varieties", containing low-contact L1 varieties, learner varieties, pidgins and creoles, or "varieties resistant to standardization" (cf. Schreier 2012).[309] Such a modified version of Modiano's model is presented in Figure 7.3.2.

The revised model is able to recognize degrees of (dis-)similarity as proposed by Sharma (2009: 173), while it goes without saying that alternative means of analysis and representation (e.g. those applied in the present study such as NeighborNets) are more adequate for quantifying these (dis-)similarities. There is no reason to suppose that other grammatical (and likewise phonological, pragmatic, etc.) features should not be localizable with the help of this model. At the same time, the weaknesses of symmetrical representations (e.g. Venn diagrams) as criticized by Hundt (2009: 305) can be avoided.

[309] Alternatively, the categorization into hyper-central, super-central, central and peripheral varieties as proposed by Mair (2013) could also be applied. This approach specifically recognizes the role of English in a global language ecology and considers both standard and non-standard varieties.

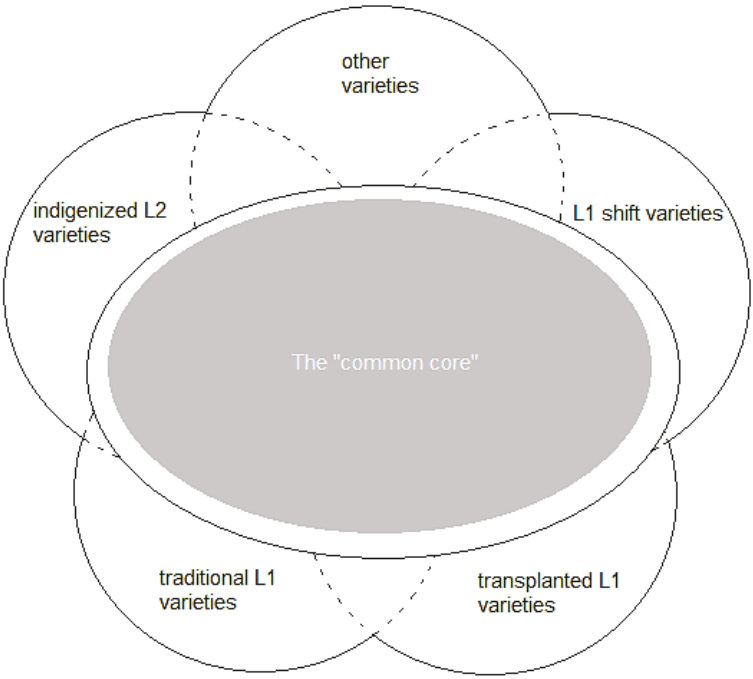


Figure 7.3.2. A revised model of Modiano's model of World Englishes: applying categories from Kortmann & Szmrecsanyi (2009)

Naturally, this is not to imply that the sociolinguistic/variety type viewpoint offered by models such as those of Kachru (and its extensions) and Schneider is meritless; it is simply less adequate for the description of linguistic structure and usage patterns. That models assessing sociolinguistic status or variety types are necessary is evident from the finding that Modiano's approach, too, cannot fully do without geographical and/or status-related circles as the underlying base for the core area. Still, there is room for improvement or adaptation. For instance, the outer circles could also be labeled according to Schneider's phases that share similar structural features (but also sociolinguistic aspects), which again illustrates the versatility of Modiano's approach (see the modified version Figure 7.3.3). The phases could be placed according to how

much they overlap, e.g. with phase 1 covering less of the core area but subsequent phases (e.g. 4 and 5) showing considerable overlap with the core area and each other.

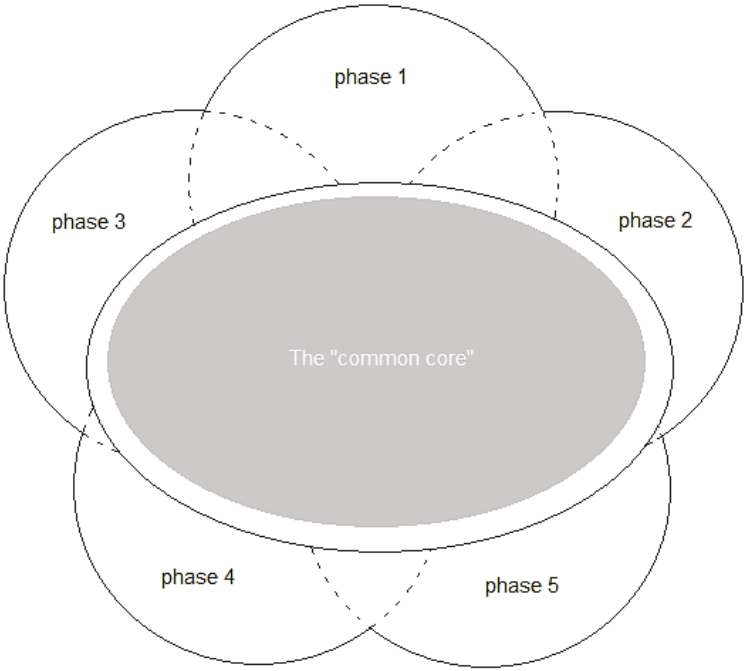


Figure 7.3.3. A revised model of Modiano's model of World Englishes: applying categories from Schneider (2007)

A further point worth discussing is the applicability of Minnich's (2004) concept of ethnoscapes (groups of people who share a form of mediated culture), which is closely related to the general phenomenon of globalization. It provides one possible explanation why globalization contributes to the blurring of linguistic boundaries, both generally and in the sense of variety types, as also hypothesized by Hundt & Vogel (2011: 161). It could be argued that the educated speakers represented in ICE constitute such an ethnoscape, though it is difficult to determine whether this is actually the case and, if so, how this ethnoscape should be de-

lineated. In order to establish a sound empirical basis for the existence of ethnoscapes, we would have to investigate the cultural habits of all contributors to the corpora and relate them accordingly in some kind of network representation. While such a project is conceivable in principle, it would require considerable resources and great methodological care by the researchers involved. For the ICE data, this is objectively beyond reach. Therefore, even if Minnich's model succeeds in relating the individual speaker perspective to general cultural developments and provides one potential point of departure that explains the overall homogeneity of varieties in the present and other areas of research, for the time being her conceptualization has to remain somewhat vague and speculative. In addition, it would be too simplistic to claim that cultural similarity necessarily implies structural linguistic similarity. The ethnoscapes model would thus seem less adequate for the study of individual structural features than, for example, Modiano's approach.

7.4 Central perfect environments and Miller revisited

Van Rooy (2009: 311) has stated that there is a lack of comparative studies that "determine potential core attributes of the perfect". In order to complement van Rooy's own attempt to tackle this issue, in this section I identify those environments where PrPfs characteristically occur. This is done in two ways. First, on the basis of findings from the above sections on the individual varieties, I consider the prototypical language-internal environment; that is, which of the factors investigated above yield high relative values and compare the findings to results from general-language studies. Second, on the basis of the PrPf-friendliness scores found above, I also establish the typical environment according to language-external categories. In other words, I look at the various text types, genres, etc. again and relate the present findings to further research. I also address Miller's (2004a: 231) statement that "the classic adverb-less PrPf exists only in formal written English" in light of the present data.

Table 7.4.1 presents an overview of the ranking of the individual language-internal variables according to their relative frequencies both for each variety and overall, which is evaluated below column by column.

The exact relative proportions for the individual variables are given when necessary.

Table 7.4.1. Ranking of language-internal PrPf variables

	temporal adverbials	Aktionsart	sentence type
GB	SP > TP > SQ > FR	ach > act = sta > acc	pos > neg > ques
IRL	SP > TP > SQ > FR	act > ach > sta > acc	pos > neg > ques
AUS	SP > TP > FR > SQ	act = ach > sta > acc	pos > neg > ques
NZ	SP > TP > SQ > FR	ach > act > sta > acc	pos > neg > ques
CAN	SP > TP > FR > SQ	ach > act > sta > acc	pos > neg > ques
PHI	SP > TP > SQ > FR	act > ach > sta > acc	pos > neg > ques
JA	SP > TP > SQ > FR	act > ach > sta > acc	pos > neg > ques
SIN	SP > TP > FR > SQ	ach > act > sta > acc	pos > neg > ques
HK	SP > TP > SQ > FR	ach > act > sta > acc	pos > neg > ques
IND	SP > TP > SQ > FR	ach > act > sta > acc	pos > neg > ques
overall	*SP > TP > SQ > FR*	*ach > act > sta > acc*	*pos > neg > ques*

	semantics	preceding TR
GB	indef-s > cont-s > indef-m > cont-c	PrTs > SPst = SPrPf > F = PrTp
IRL	indef-s > cont-s > indef-m > cont-c	PrTs > SPst > SPrPf > F > PrTp
AUS	indef-s > indef-m = cont-c = cont-s	PrTs > SPrPf > SPst > PrTp > F
NZ	indef-s > cont-s > cont-c > indef-m	PrTs > SPrPf > SPst > PrTp = F
CAN	indef-s > indef-m = cont-c > cont-s	PrTs > SPst = SPrPf > PrTp > F
PHI	indef-s > cont-s = cont-c > indef-m	PrTs > SPst = SPrPf > F > PrTp
JA	indef-s > cont-c > cont-s > indef-m	PrTs > SPrPf > SPst > PrTp > F
SIN	indef-s > cont-c > cont-s > indef-m	PrTs > SPst > SPrPf > F > PrTp
HK	indef-s > indef-m > cont-c = cont-s	PrTs > SPrPf > SPst > F > PrTp
IND	indef-s > cont-s > cont-c = indef-m	PrTs > SPrPf > SPst > PrTp > F
overall	*indef-s > cont-s > cont-c > indef-m*	*PrTs > SPrPf > SPst > F > PrTp*

If a PrPf is temporally specified, as is the case for 23.2% of all PrPfs in the data from the ICE samples, this is typically done by way of adverbials of span and duration (SP; 46.4%) or those referring to a point in time (TP; 30.1%). The remaining two categories, adverbials of sequence (SQ; 13.6%) and adverbials of frequency (FR; 9.9%), are much less common. However, the hierarchy SP > TP > SQ > FR is not observable in all varieties. While SP > TP (> FR/SQ) seems to represent a stable pattern, the positions of SQ and FR vary. In the sections on the individual varieties above, it was hypothesized that the high percentage of SQ adverbials (e.g. *already*) in some of the L2 varieties (PhiE, JamE, HKE, IndE) may

be due to their use as additional markers of completion, most likely as a result of substrate influence. In fact, they are also ranked third in BrE, IrE and NZE. However, the hypothesis can still be considered valid as mode of discourse seems to play a part here. In the three Asian L2 varieties (especially PhiE and HKE, but also IndE), SQ adverbials are particularly prominent (20% or more of all specified tokens) in speech. This finding strongly suggests that, even in acrolectal/upper-mesolectal English as contained in the ICE data, we can observe substrate transfer not only on a more superficial level (e.g. with regard to lexical items, characteristic discourse markers or alternative perfect forms) but also with respect to frequencies of individual variables for one particular area of grammar (in our case the PrPf). Speech, of course, is the mode of discourse where this is most likely to occur. This notion dovetails with the finding that SQ adverbials are overrepresented in involved texts (see Section 6.3).

In the ranking of Aktionsart categories, another split picture emerges. Achievements (39.2%) overall constitute the majority class, closely followed by activity verbs (37.8%). Minority variants are state verbs (20.2%) and the rare category of accomplishments (2.8%). Even if the relative frequencies of achievements and activities do not diverge by a large margin, achievements seem to be overrepresented in PrPfs provided that verbs of the activity type are commonly viewed as the most frequent category overall (covering nearly 50% of all verbs; cf. Biber et al. 1999: 365). This in turn strengthens the position of achievements as the central Aktionsart class with PrPfs when considering the aggregate data. Again, the ranking of the two highly frequent classes is not fully homogeneous across the varieties; some subtleties are worth closer inspection. Activity verbs are the most common class in a number of varieties (IrE, AusE, PhiE, JamE), while it could be argued that among the L2 varieties achievements are most prominent in the Asian ones (SinE, HKE, IndE), two of which are still oriented toward BrE. Note further that state verbs are nearly as frequent as activities in BrE, which is therefore the only variety showing such a pattern. This latter finding contradicts Elsness (1997: 321), whose results suggested a low frequency of statives with the PrPf. In addition, the present findings for the ranking of Aktionsart categories (achievements > activities > states > accomplishments) can be

related to Liszka's (2003: 18) hypothesis that "adequate usage" of the PrPf by learners of English is "in descending order of the following hierarchy of Aktionsart categories: states > activities > accomplishments > achievements". There is some limited evidence in the ICE data that corroborates this hypothesis. When we consider student writing (w1a) on its own, the JamE and IndE components support the view of a comparatively extended usage of PrPfs with states. However, in the remaining L2 varieties, this seems not to be the case; thus, Liszka's claim cannot be universally confirmed.

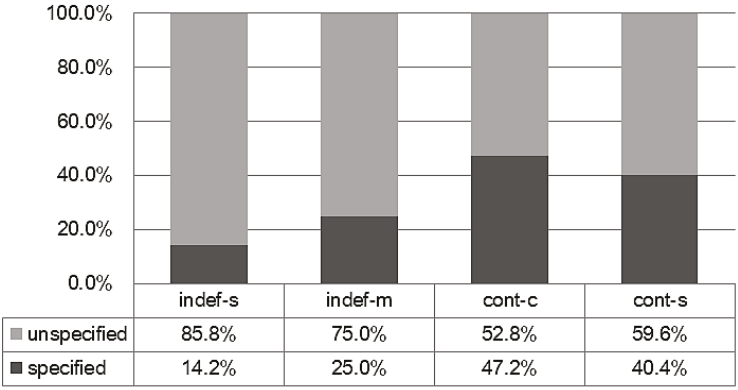
PrPfs across the board most often (89.4%) occur in positive statements, while they are less frequent in negatives (9.3%) and least common in questions (1.3%). This large bias toward the first sentence type is unlikely to be merely due to the presence of a PrPf but to the prevalence of assertive discourse in general. Negative discourse, in contrast, is more restricted in its communicative scope and thus used less often than assertive discourse (< 10%; cf. Biber et al. 1999: 1115; Mindt 2000: 158). Genre effects play a part here, however, and Schlüter's finding that PrPfs in negatives (17.1%) and questions (14.0%) are noticeably more frequent in conversation (2002a: 135–136)[310] can be extended to further varieties of English as well as to writing where conversation is simulated, as was the case in a number of creative writing (w2f) texts and occasionally in letters (w1b). Still, the genre values (dialogues) for positives do not fall below 75% in any of the ICE samples, indicating that this sentence type is a central environment for the PrPf.

The vast majority of PrPf occurrences (62.4%) lend themselves to an indefinite single act reading, while the shares of the remaining categories are fairly equal. In detail, continuative state readings (13.9%) are slightly more frequent than continuative acts (12.6%) and indefinite multiple acts (11.1%). While the percentage for the indefinite single reading almost exactly matches that of Schlüter (2002a: 166) for BrE and AmE (62.8%), the remaining frequencies diverge and some variability

[310] Although no directly comparable percentages are available from Biber et al. (1999: 159, 211), the substantially higher rate of negatives and questions in conversation in general is apparent from the normalized frequencies (per million words) provided (questions: 0.0235 in conversation, 0.0070 in fiction, 0.0005 in news and academic writing; negatives: 0.0220 in conversation, 0.0135 in fiction, 0.0065 in news, 0.0005 in academic writing).

can be observed in the individual varieties. When ranking the frequencies of the semantic readings, indefinite single act readings are the majority variant throughout, while the remaining categories can be found on all following ranks, apparently leaving us with an inconclusive result. Accordingly, on the basis of the present data, I will not attempt a comprehensive definition of the *basic* meaning of the PrPf (in the sense of Klein & Vater 1998: 216; see also sections 3.3 and 4.3.4). The lowest common denominator for all four readings would be a broad interpretation such as "reference to some kind of anteriority" or "once" (see also Portner 2003: 492). This would remain unsatisfactory as (i) such a definition is not exclusive to the PrPf (as it would also apply to other TR forms creating past time reference) and (ii) it would fall short of including less frequent semantic readings such as continuative or specialized facets of meaning such as hot-news/recent past, etc. (see Section 3.3.2). However, what we can establish is the *central* meaning (see Klein & Vater 1998: 216) of the construction for worldwide varieties of English, which is "reference to a single event or situation in the indefinite past". This definition is in accordance with and broadens the results of a corpus-based study of BrE/AmE (Schlüter 2002a: 167), whose model has proven viable for the current analysis. In addition, it has the advantage that it still allows for further readings to be included and is thus more apt for the semantic versatility of the PrPf. That indefinite single act constitutes the most central meaning can also be deduced from the fact that the remaining readings – the continuative types in particular – yield higher proportions of temporal specification and therefore often need some kind of contextual support, while the indefinite types are more context-independent (see Figure 7.4.1).[311]

[311] Differences between the indefinite (indef-s + indef-m) and the continuative group (cont-c + cont-s) are highly significant $\chi 2 = 452.35$, df = 1, $p < 0.001$, $\varphi_C = 0.291$). This result from the ICE samples again broadens Schlüter (2002a: 260) beyond the BrE/AmE paradigm. Note besides that in contrast no evidence for a correlation between Aktionsart categories and semantic readings could be found in the data.



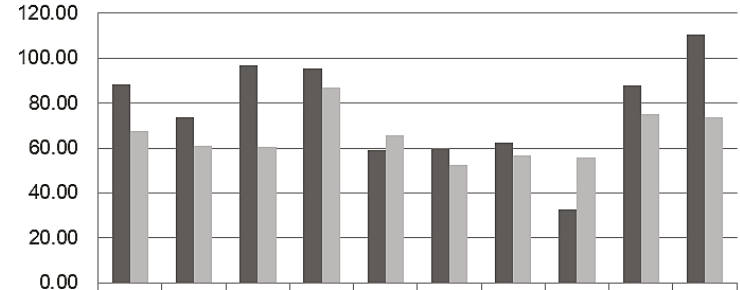
	indef-s	indef-m	cont-c	cont-s
unspecified	85.8%	75.0%	52.8%	59.6%
specified	14.2%	25.0%	47.2%	40.4%

Figure 7.4.1. Relative proportions of PrPfs with and without temporal specification in ICE: semantic categories

The final point to consider from a language-internal perspective is the ranking of preceding TR forms. PrTs is the most common variant (57.8%), while all other forms – apart from those that create past time reference, SPrPf (17.8%) and SPst (16.0%) – are negligible (< 4%). Note that the ranking of the latter two forms differs across varieties. It is not surprising that PrTs is the most common preceding TR form, as present tense verbs have been found to be the most frequent category in English in general. However, mode of discourse again is important. PrTs forms seem to be underrepresented in PrPf contexts in particular when compared to general-language frequencies (Biber et al. 1999: 456) for conversation (> 67%) and academic writing (> 66%), while overrepresented compared to fiction (40%) and news (52%). Still, it could be argued that the PrPf, due to the frequent present tense contexts in which it is embedded, should be seen as a present tense itself or at least as a form that strongly suggests current relevance or "presentness" (see also below) and is therefore likely to be triggered by preceding PrTs constructions. Again, because this may merely be the central context in which it appears, other options are available, as shown by the strong presence of the past TR forms (SPrPf and SPst) as priming factors. Taking into account the overrepresentation of present tense forms mentioned above, the results nicely illustrate the hybrid nature of the PrPf, which may surface

in both present and past time contexts (including – birds of a feather flock together – occurrences with other PrPfs).

I will now establish a second dimension of typicality of the PrPf according to language-external criteria, that is, textual categories. However, I will not present an in-depth comparison of the PrPf-friendliness of varieties with each other or discuss issues such as the alleged Americanization of PrPf usage (that is, declining PrPf frequencies) in some varieties, given that (i) these topics have been expertly covered elsewhere (see Yao & Collins 2012, Hundt & Biewer 2007 or Sand 2005: 457) and (ii) the ICE data do not allow for the establishment of diachronic trends.



	GB	IRL	AUS	NZ	CAN	PHI	JA	SIN	HK	IND
spoken	88.42	73.83	96.93	95.64	59.01	59.60	62.57	32.48	87.88	110.56
written	67.40	60.97	60.52	86.89	65.75	52.42	57.00	55.76	75.07	73.73

Figure 7.4.2. Perfect friendliness of varieties (normalized frequency per 10,000 words)

However, it is worth briefly noting that neither the current data (see Figure 7.4.2) nor the results from Yao & Collins (2012: 392) support Bautista's (2008: 202, 217) claim of a cline of closeness of Asian varieties in relation to "norms of the Inner Circle" in terms of PrPf-friendliness along the lines BrE > SinE > PhiE > HKE.

Figure 7.4.2 further shows that in the majority (8/10) of the ICE samples, PrPfs are more frequent in speech (overall average = 76.69; $\sigma = 22.24$) than in writing (overall average = 65.55; $\sigma = 10.02$),[312] which is

[312] A similar result emerges when we re-group the data according to involved versus informational texts. Involved texts (74.81) are on average more PrPf-friendly than the infor-

the more internally homogeneous mode of discourse. Exceptions to this pattern are CanE and SinE, two varieties whose PrPf-friendliness is clearly below average across the data. Note that the low value for speech in the latter is due to an extremely low count for one text type, private dialogues (s1a: 7.60 PrPfs/pttw). However, the values for the non-scripted spoken categories in the ICE-SIN data are also particularly low (s1b: 22.50; s2a: 27.00) in contrast to the scripted spoken category (s2b: 72.80). Interestingly, comparable patterns (that is, considerably higher PrPf-friendliness in scripted monologues compared to the non-scripted spoken texts), although with less extreme contrasts, are in evidence in all varieties under investigation, apart from NZE. A possible explanation for this finding is that news discourse, as represented in s2b, naturally relies on PrPfs to locate a news item in the (indefinite) past (see also Yao & Collins 2012: 392).[313] SinE therefore seems to have the most marked contrast between scripted and natural spoken discourse (which relies on other items for creating past time reference, such as temporal adverbials; see the examples in Section 5.8).

For CanE, another explanation for its lower PrPf friendliness in speech is conceivable. It could be claimed that the salience of the "colloquial preterite" (Vanneck 1958; cited in Elsness 2009a: 92) is responsible for the lower PrPf frequencies in speech and dialogues in this variety in particular. However, one has to be cautious when establishing such a connection, as shown by studies of the PrPf-SPst alternation (e.g. Wynne 2000: 45 or Schneider & Hundt 2012: 21).

Given the results outlined above and taking account of the exceptional patterns just mentioned, it comes as no surprise that, overall, the macro-genre monologues has the highest PrPf-friendliness (average across the data = 86.04; σ = 19.89). However, scripted monologues (s2b: 100.60; σ = 18.92) do not show the highest average overall; this position is taken instead by persuasive writing (w2e: 111.80; σ = 28.95) in the cross-varietal perspective. It has to be acknowledged, though, that the

mational texts (63.72), while HKE and SinE are the only varieties where the proportions are reverse.

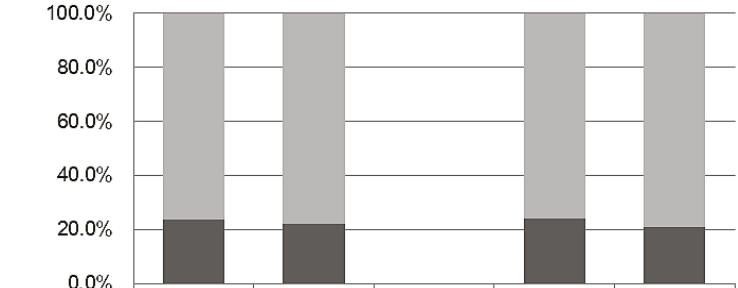
[313] Note in addition that s2b, as a scripted category, is a likely candidate to appear removed from the remaining spoken types (Biber & Conrad 2009: 261). This is indeed the case in four out of the ten varieties (BrE, NZE, PhiE, IndE).

dispersion is much more pronounced in the latter. Note also that letters (w1b: 94.50; $\sigma = 25.78$), another textual category that could be interpreted as a monologic scripted type (though dialogic elements are undoubtedly present), reaches the third highest PrPf-friendliness value, and reportage (w2c: 79.15; $\sigma = 7.44$) ranks fourth.[314] The high value for letters (see also Van Herk 2008: 51) can be interpreted as a corroboration of Wright's (1986: 84) hypothesis that PrPfs are highly frequent in this genre as they serve to convey "structural subjectivity", although the notion itself admittedly remains vague. A potential solution for making the structural subjectivity of text types more visible with the help of quantitative corpus data would be to measure the salience of indexicals (items such as *here* or first- or second-person pronouns; see Guenthner 1977: 94–95 or Weinrich 2001: 208–213) as markers of "presentness" (Bowie et al. 2013: 349) or subjective discourse in texts of a given type, and to correlate these values with the respective PrPf-friendliness scores.

As indicated above, Miller maintains that "the classic adverb-less PrPf exists only in formal written English" (2004a: 231). While this might be true for BrE/AmE, the ICE data do not support this claim. Figure 7.4.3 below reveals that PrPfs in speech are more often temporally specified by an adverbial than PrPfs in writing. Differences between the two modes of discourse are not significant, and still more than three quarters of all occurrences are without specification. Even when we regroup the data into involved versus informational texts in order to better meet Miller's "formal" (~informational) condition, the proportion of PrPfs that are not specified still exceeds three quarters. The two categories are significantly different, although with only a very weak effect ($\chi^2 = 4.76$, $df = 1$, $p < 0.05$, $\varphi_C = 0.03$). Thus, Miller's claim would seem to be overstated in light of the current data; an improved version that could

[314] Reportage and scripted monologues are fused into a "news" category in Yao & Collins (2012). This category emerges as the most PrPf-friendly in nine out of ten varieties in their study, which ties in with the present results. Another parallel is that creative writing (w2f: 33.30; $\sigma = 16.04$) represents the least PrPf-friendly text type, a result that is also in line with Van Herk (2008: 51). A comparison to Biber et al. (1999: 461–462; cf. Section 3.1 above) yields similarities and differences. While they also identify "news" as a PrPf-friendly category in their BrE/AmE data, the results from studies including further varieties (the present study included) contradict Biber et al.'s identification of fiction as another PrPf-friendly type.

be considered valid beyond the BrE/AmE paradigm could be reformulated as follows: "The classic adverb-less PrPf is *more likely* to be found in formal written English".



	spoken	written		involved	informational
unspecified	76.2%	77.8%		76.1%	78.9%
specified	23.8%	22.2%		23.9%	21.1%

Figure 7.4.3. Relative proportions of PrPfs with and without temporal specification in ICE: spoken vs. written and involved vs. informational

8 Conclusion and outlook

This study offered an account of the PrPf and related aspects in worldwide varieties of English, and can be seen as another stepping stone toward a comprehensive understanding of the many uncertainties that surround this particular construction and its usage patterns. It relied on ample, "tidy" corpus data from the ICE project, providing a much larger statistical base than previous empirical studies of the subject. These corpus data were analyzed using both a quantitative and a qualitative approach. The former approach was taken mainly in the sections on variation in PrPf usage across as well as within individual varieties, registers, text types, etc., while the latter was predominantly (though not exclusively) applied in the case studies. The study allowed us to follow up on two desiderata in the area of PrPf research, namely (i) to extend our knowledge of this grammatical area beyond the traditional BrE/AmE paradigm, and (ii) to take account of language-external effects that as yet have received only scarce attention. For instance, the study considers intravarietal variation via a systematic comparison of modes of discourse, macro-genres and text categories.

Although a number of contextualizations and generalizations were already given in the discussion section (see Chapter 7), this concluding section summarizes the main findings and addresses the research questions that were posed at the outset of the study (see Chapter 1). In addition, it contains some reflections on methodological aspects and sketches potential avenues for further research that emerged in the course of the present investigation.

After addressing the history of the English PrPf, outlining its status as described in the relevant theoretical literature, and reviewing the related corpus-based publications, the study presented a fine-grained analysis along language-internal and language-external variables in individual varieties. These sections provided a wealth of quantitative data that, on the one hand, corroborated earlier research with a similar focus but with a restricted set of varieties, and, on the other hand, may be used for future reference. Possible applications include the creation of corpus-based grammars of the individual World Englishes (or indeed a single

grammar of World English(es)), which are yet to be compiled. As could be expected, variation was observed with respect to PrPf-friendliness values and the relative distributions of the language-internal variables. However, the most striking results were as follows: (i) spoken text types aligned closely across nearly all varieties (except SinE), yet to date mode of discourse has not been considered influential in studies of tense and aspect, and (ii) this result was consistently observed despite the fact that we only considered the contextual and inherent factors of one particular grammatical feature. When single factors were analyzed for individual varieties, the written text types more often than not yielded comparatively more variability in terms of distribution, again testifying to the greater variability of this mode of discourse. Further, the extensive number of temporally specified PrPfs alleged to exist in Asian varieties, notably IndE and HKE, could not be traced in the ICE data. A quantitative analysis involving alternative perfect forms might yield a different result; however, an investigation of this kind would require different, ideally more spoken informal and basilectal, data.

The application of an aggregational multidimensional approach to the corpora yielded a number of insights. Above all, PrPf usage emerged as a globalized and core feature across World Englishes, albeit with two qualifications. First, the input data represents the language of a restricted set of speakers only (educated speakers in accordance with the principles of ICE). However, the advantages of working with a well-balanced and relatively clean set of language material outweigh the possible disadvantages, such as a disproportional amount of effort in the data compilation stage. Therefore, we accepted this trade-off as it is not detrimental to the overall meaningfulness of the analysis. Second, the analyses succeeded in showing that, while generally there is only scarce evidence of the explicit nativization of PrPf usages (with regard to overall differences of distributions of the language-internal factors), there are fine nuances that distinguish the varieties under investigation. All of this was neatly represented in the NeighborNet (Figure 6.2 in Section 6.1) that subsumed the overall relationship between all varieties as roughly equidistant from one another and therefore could be described as representing "sameness in differences". The sections on the individual varieties offered possible explanations for the variation, such as learner ef-

fects and substrate influence (or both phenomena reinforcing one another), particularly in multilingual settings.

These caveats surface both in the multidimensional studies of language-external variables across all varieties and in the case studies, albeit qualitatively. In the former, geolinguistic signals or groupings according to variety types were largely absent, but overlap was to a certain extent determined by text type. A significant split emerged when texts were grouped according to the dimension involved versus informational, as established by Biber (1988). This is indicative of the applicability and value of Biber's approach beyond traditional register studies. It also further emphasizes the importance of studying register variation in the first place in order to account for internal variability in the data, as Sand (2005: 458) and Schneider & Hundt (2012: 29–30), for example, have proposed for the study for World Englishes. Taking the opportunity offered by the parallel design of the ICE components proved worthwhile; the analyses showed that occasionally comparable genres or text types of different varieties are closer to one another than associations within varieties or varieties of the same type. This supports the notion of exploring linguistic realities rather than simply conceiving of (regional) varieties as monolithic blocks. For the feature under scrutiny in this study, this internal variability was restricted, but increased markedly the more fine-grained the analyses became. In addition, the register analysis revealed that PrPfs were more frequent in speech (particularly in monologues) in nearly all varieties, and that there were strong text type effects on PrPf friendliness (texts with high values were scripted monologues, letters and persuasive writing).

Creative usage patterns were observed in the data. In the case studies, non-standard surface forms (that is, forms other than HAVE + V-*en*) in perfect contexts were restricted to spoken informal genres. Their spread across varieties of all types suggests that these forms could be seen as true angloversals or, alternatively, as (spoken) "register versals". However, further specialized semantic or pragmatic values were not found. The second case study revealed that temporal adverbials that were considered typical in perfect contexts in fact allowed for variation to different extents or even correlated more strongly with the Simple Past, which suggested layering. In this area, a split between L1 and L2 varie-

ties was in evidence, while differences between the traditional L1, BrE, and the transplanted L1s were less marked and Schneider's (2007) phase model could be fruitfully applied for categorizations in this particular area. Furthermore, the third case study showed that PrPf usage in definite temporal contexts occurs in the ICE data, albeit rarely and predominantly in speech, the crucial point being that clear qualitative differences between L1 and L2 varieties existed. However, the case for a current or future development of the PrPf into a "proper" tense (in the sense of Quirk et al. 1985) was found to be weak at best.

In light of the findings from the empirical study, some of the models in World Englishes were critically reviewed. It was established that approaches such as Modiano's model possess more explanatory potential when investigating individual structural features of and across varieties than more traditional approaches (e.g. Kachru's circles). Schneider's recent dynamic model (and, to a lesser extent, also Kachru's approach) still proved useful in accounting for structural differences between varieties, even if those take a genuinely sociolinguistic-historical perspective. Other models (e.g. by Minnich) provided interesting approaches toward incorporating aspects of globalization but remained vague in their description. For these reasons they have not received wider acceptance in the research community. In sum, I argued that the various models should rather be seen as complementary parts, and that analyses of features of World Englishes need to take account of findings from language contact studies to achieve a sound overall picture (see Schreier & Hundt 2013). Accordingly, two revised versions of Modiano's model, which incorporate categorizations of the sociolinguistic approaches, were sketched.

Another result that transpired, albeit one with less theoretical impact, from the analysis of the ICE data was that the prototypical or central PrPf in World Englishes occurred

- without temporal specification (if it is specified, specification is most likely by an adverbial of span and duration);
- with an achievement verb;
- in a positive/assertive sentence;

- in a semantic context which suggests an indefinite single act reading;
- after a Simple Present.

However, contrary to what this reductive list suggests, the analyses also showed that PrPfs surface in combination with a wide range of contextual variables (see Section 7.4). For instance, in the area of Aktionsart, activity verbs were nearly as frequent as achievements. Variation in patterns also occurred across genres, exemplified by the higher frequencies of negatives and questions in conversation. That the PrPf covers a special position in the system of TR forms in the World Englishes under investigation could be deduced from the finding that it was temporally specified much more often than other TR forms, in particular when non-central semantic interpretations that required contextual support seemed likely. In addition, the PrPf emerged as a hybrid form that was salient not only in present tense environments, but also in reference to the past (either by preceding SPst or PrPf). Therefore, it should be viewed as being characterized by great versatility, and given this inherent complexity, we would do well to speak of central contexts, recognizing that more peripheral ones also exist.

Once again, the ICE data facilitated the incorporation of register as a language-external predictor. In this perspective, the PrPf was established as a form that in World Englishes typically appears in scripted monologic speech. It goes without saying, however, that this view is overly simplistic due to the multifaceted nature of the corpus data. It is crucial to keep in mind that this context represented merely the central language-external environment and that PrPf frequencies may yield considerable variability when more fine-grained categories are applied, as shown in the case of persuasive writing (w2e), the single most PrPf-friendly text type.

While the establishment of the central PrPf contexts represents an interesting finding in itself and the aim of this study is not principally pedagogical, the practical relevance of some of the results should not be discounted. First, the central contexts established could serve as the basis for a reassessment of the taxonomies or curricula that are presently

employed in teaching the Present Perfect. Learners of English at a basic stage, who are confronted with the construction for the first time, should, as an initial step, be familiarized with its typical (= central) uses and environments. To briefly outline a concrete example, teaching should start out with model sentences involving indefinite single act interpretations (involving an achievement verb, etc.), as learners are much more likely to encounter this usage in their real-life contacts with speakers of English (of whatever variety type) or when dealing with textual material (newspapers, books, blogs, etc. from any variety). Model sentences and real-life material involving less central context variables should be introduced at a more advanced stage. Another pedagogical aspect relates to one of the issues referred to by Wynne (2000), namely that many textbooks and usage guides use lists of temporal adverbials that allegedly act as triggers for the PrPf. The present study suggests that the inclusion of some of the items on these lists (e.g. *ever* and *recently*) should be reconsidered, as the corpus data did not provide unequivocal evidence for an extended co-occurrence of these items with the PrPf in either the L1 or L2 Englishes. In that respect, collostructional analyses, for example on associations between items such as *ever* with either the PrPf or the SPst in particular lexical environments (*Have you ever* been *to Russia?* vs. *Did you ever* go *to Russia?*), could serve to reveal further intricacies in the variation between the two TR forms.

The following points are worth summing up from a methodological point of view. On the one hand, the study indicated that multidimensional aggregational methods such as cluster analyses and phylogenetic networks were helpful tools for establishing underlying structures in large sets of data and can help to reduce complexity in insightful ways. On the other hand, even more than 5,000 data points (or 6,000 in the case study on temporal adverbials) sometimes produced low individual frequencies for a particular variable (or item in the case study), textual category or variety. More data points would occasionally have been desirable, in particular to enable us to achieve a better fit between the numerical data and the graphical output. The aggregational quantitative analyses were complemented by distributional tables and a qualitative approach (e.g. the study of more than 300 individual examples) to areas where the data required in-depth analysis. Given a large enough (but not

excessively large) dataset, the combination of different analytical techniques used in the present study helped to achieve a sound overall picture and to arrive at meaningful conclusions.

However, a short note on the problem of data scarcity is in order here. This was especially salient in sections such as the case studies on alternative perfect forms or the usage of the PrPf as a narrative tense, where sources other than ICE (such as the recent GloWbE corpus; Davies 2013) would likely have yielded some quantifiable insights into the phenomena. Thus, for future research on these topics two things would be desirable:

(i) For a quantitative assessment of the frequency and usage patterns of alternative surface forms, complementing the corpus-based approach (ideally including mesolectal and basilectal data) with elicitation tests or acceptability questionnaires would be indispensable (see Krug & Sell 2013). Survey data of this kind from speakers of different varieties (potentially also controlling for different registers) could reveal further intricacies in these genre/register effects, and could help to explain the scope of the convergence of the varieties in this respect.

(ii) To examine the case for the existence of alternative forms as vernacular *uni*versals, additional comparisons across more languages would be necessary. With reference to the case study on the usage of the PrPf as a narrative tense, further work, for example involving longitudinal quantitative study of both L1 and L2 varieties, is needed to establish whether the phenomenon of PrPf usage with definite temporal adverbials has been diachronically spreading or is rather less regular and may therefore be attributed to learner and substrate effects alone.

Related aspects worth investigating are when exactly the PrPf emerged as a globalized or core feature of worldwide varieties of English, and whether the suspension of the temporal adverbials constraint is grounded in historical input varieties or rather represents a truly innovative

feature. This would require historical data from earlier stages of colonization, which, particularly for most of the L2 varieties, are scarce. Relevant corpus projects have been initiated only recently.[315]

Note that genuine learner varieties, such as those investigated by Bardovi-Harlig (2000) and represented by the ICLE family of corpora (Granger et al. 2009), are not included in the present study but represent an interesting area of further research (e.g. a comparison of corpus data from ICE and ICLE). The same applies to comparisons with material from varieties of English not included in ICE, such as Jersey English (Jersey Interview Corpus/Jersey Archive Corpus; see Rosen 2014). However, as ICE, ICLE and further corpora are not directly comparable in respect of their textual categories, evaluations would have to remain on a more global level. In contrast, broadening the scope of the findings by including data from the ever-growing family of ICE components (e.g. for Sri Lanka, Nigeria or Malta) would also allow the investigation of register effects. Extending the analysis to material from ICE-USA (see Werner in preparation b) in particular would allow us (i) to determine the strength of associations between postcolonial varieties and their major donor varieties (BrE/AmE), or at least (ii) to better quantify the extent of Americanization of present-day varieties. Assuming that Americanization can, to a certain degree, be equated with globalization, this would enable us to test whether (linguistic) globalization can be traced even in the distributions of particular features (see Krug & Rosen 2012 for a study on lexical items). However, Hickey (2004: 3) has emphasized that it might be difficult to establish such connections when highly pervasive features are investigated – and the PrPf and its usage clearly qualify as such.

Using the ICE data, it also appears to be possible to conduct comparable aggregational "inner life" analyses for structural features other than the PrPf or for combinations of features, which would provide an alternative to the global approach (mere presence/absence or pervasiveness ratings) as provided in Szmrecsanyi & Kortmann (2009) or Kortmann & Lunkenheimer (2011). A related point where further work is indispensable are comparative studies within a similar framework as the present one, establishing a grammar of usage for other "rival" TR forms,

[315] See also the discussion in Mukherjee & Schilk (2012: 195–196).

notably the PrTs and the SPst (see Davydova 2011 or Yao & Collins 2012, for instance). Although such projects would naturally involve an extensive amount of manual coding and analysis, they would enable us to relate PrPf patterns and context variables to those of the other TR forms directly. Consequently, this would allow us to go beyond the fairly general assessments of the distributions of the context variables as provided in Biber et al. (1999), for example (see Chapter 7.4), and facilitate the descriptive analysis of specific usage patterns of these forms.

In sum, again with particular reference to the considerable overlap that exists between varieties, the findings of this study can be located in the (fledgling) tradition of corpus-based studies of individual features (and related usage patterns) that emphasize the convergence of World Englishes as regards regular(izing) patterns, while still taking account of related, less frequent occurrences. Despite having investigated one grammatical feature only, I hope to have shown that global phenomena are of an inherently complex and multifaceted nature. Overall, worldwide varieties of English show a considerable degree of convergence or uniformity. Still, variation – be it in the form of quantitative differences for a number of variables or qualitative differences as presented in the case studies – is clearly evident, lending each variety under investigation a unique set of properties. This is the case despite the fact that only educated (and arguably mediated) language as contained in ICE was included in the analysis. Linguistic realities are thus, at times unexpectedly, far removed from what the neat models suggest and we have to accept that a truism of social study, namely that "global things are seldom as orderly as we wish they were" (Elliott 2010: xv), applies equally to linguistic analysis. However, to end on a positive note, this actuality also leaves ample opportunity for researchers of World Englishes to describe and explain the complexity of variation in the future.

Appendices

Appendix A. ICE text categories
(respective number of 2,000-word texts indicated in brackets)

Register/ mode of discourse	Macro-genre	Text type category	Text type (detailed)
Spoken (300)	Dialogues (180)	s1a Private (100)	Face-to-face conversations (90) Phone calls (10)
		s1b Public (80)	Classroom lessons (20) Broadcast discussions (20) Broadcast interviews (10) Parliamentary debates (10) Legal cross-examinations (10) Business transactions (10)
	Monologues (120)	s2a Unscripted (70)	Spontaneous commentaries (20) Unscripted speeches (30) Demonstrations (10) Legal presentations (10)
		s2b Scripted (50)	Broadcast News (20) Broadcast Talks (20) Non-broadcast Talks (10)
Written (200)	Non-printed (50)	w1a Student writing (20)	Student essays (10) Exam scripts (10)
		w1b Letters (30)	Social letters (15) Business letters (15)
	Printed (150)	w2a Academic writing (40)	Humanities (10) Social sciences (10) Natural sciences (10) Technology (10)
		w2b Popular writing (40)	Humanities (10) Social sciences (10) Natural sciences (10) Technology (10)
		w2c Reportage (20)	Press news reports (20)
		w2d Instructional writing (20)	Administrative writing (10) Skills/hobbies (10)
		w2e Persuasive writing (10)	Press editorials (10)
		w2f Creative writing (20)	Novels and short stories (20)

Appendix B. Relative values of factors: text categories involved vs. informational (large percentage differences (> ±4%) highlighted)

factor	AVG INV	AVG INF	DIFFERENCE
with TA	23.43%	20.80%	2.63%
without TA	76.57%	79.20%	-2.63%
TP	36.10%	24.08%	12.02%
SP	41.39%	52.73%	-11.34%
FR	8.42%	7.59%	0.83%
SQ	14.09%	7.27%	6.83%
state	20.59%	19.50%	1.09%
activity	36.25%	40.49%	-4.25%
accomplishment	3.09%	3.15%	-0.06%
achievement	40.00%	36.85%	3.14%
pos	89.97%	91.29%	-1.32%
neg	8.92%	7.70%	1.22%
ques	1.11%	1.01%	0.10%
indef-s	64.70%	58.27%	6.43%
indef-m	10.77%	14.54%	-3.77%
cont-c	12.17%	12.63%	-0.46%
cont-s	12.36%	14.56%	-2.21%
PrTs	60.23%	55.98%	4.25%
PrTp	3.54%	2.80%	0.74%
SPst	13.34%	20.97%	-7.63%
Pstp	0.45%	0.19%	0.25%
SPrPf	17.73%	14.81%	2.92%
PrPfp	0.51%	0.74%	-0.23%
PaPfs	0.43%	0.58%	-0.14%
PaPfp	0.00%	0.00%	0.00%
F	3.85%	3.94%	-0.10%

Appendix C. Absolute and relative frequencies of Present Perfect/Simple Past Tense co-occurrence for indefinite temporal adverbials

	GB				IRL				AUS			
	PrPf		SPst		PrPf		SPst		PrPf		SPst	
	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.
already	75	68.18%	35	31.82%	51	71.83%	20	28.17%	44	67.69%	21	32.31%
yet	12	63.16%	7	36.84%	8	66.67%	4	33.33%	5	55.56%	4	44.44%
always	60	48.78%	63	51.22%	42	38.53%	67	61.47%	26	49.06%	27	50.94%
ever	20	41.67%	28	58.33%	26	30.59%	59	69.41%	23	47.92%	25	52.08%
never	72	48.00%	78	52.00%	44	25.43%	129	74.57%	55	44.00%	70	56.00%
recently	16	57.14%	12	42.86%	5	33.33%	10	66.67%	3	21.43%	11	78.57%
just	65	86.67%	10	13.33%	52	82.54%	11	17.46%	76	58.02%	55	41.98%
since	38	82.61%	8	17.39%	43	78.18%	12	21.82%	28	90.32%	3	9.68%

	JA				SIN				HK			
	PrPf		SPst		PrPf		SPst		PrPf		SPst	
	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.
already	25	46.30%	29	53.70%	34	41.98%	47	58.02%	103	70.55%	43	29.45%
yet	3	50.00%	3	50.00%	0	0.00%	3	100.00%	5	71.43%	2	28.57%
always	44	44.44%	55	55.56%	20	33.33%	40	66.67%	52	47.27%	58	52.73%
ever	11	47.83%	12	52.17%	19	57.58%	14	42.42%	10	37.04%	17	62.96%
never	35	22.29%	122	77.71%	16	18.39%	71	81.61%	53	37.59%	88	62.41%
recently	11	39.29%	17	60.71%	4	21.05%	15	78.95%	21	43.75%	27	56.25%
just	25	44.64%	31	55.36%	26	26.26%	73	73.74%	68	73.91%	24	26.09%
since	24	75.00%	8	25.00%	19	73.08%	7	26.92%	41	73.21%	15	26.79%

	NZ				CAN				PHI			
	PrPf		SPst		PrPf		SPst		PrPf		SPst	
	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.
already	46	60.53%	30	39.47%	58	73.42%	21	26.58%	76	39.79%	115	60.21%
yet	6	75.00%	2	25.00%	15	78.95%	4	21.05%	2	28.57%	5	71.43%
always	76	49.35%	78	50.65%	49	42.24%	67	57.76%	80	59.26%	55	40.74%
ever	23	42.59%	31	57.41%	27	47.37%	30	52.63%	11	42.31%	15	57.69%
never	72	41.86%	100	58.14%	60	34.48%	114	65.52%	51	41.80%	71	58.20%
recently	17	39.53%	26	60.47%	9	36.00%	16	64.00%	7	26.92%	19	73.08%
just	87	75.00%	29	25.00%	45	49.45%	46	50.55%	29	36.25%	51	63.75%
since	33	91.67%	3	8.33%	35	85.37%	6	14.63%	31	77.50%	9	22.50%

	IND				EA				NIG			
	PrPf		SPst		PrPf		SPst		PrPf		SPst	
	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.	abs. freq.	rel. freq.
already	88	80.00%	22	20.00%	113	70.63%	47	29.38%	26	59.09%	18	40.91%
yet	2	28.57%	5	71.43%	8	61.54%	5	38.46%	0	0.00%	3	100.00%
always	30	48.39%	32	51.61%	65	61.90%	40	38.10%	29	43.94%	37	56.06%
ever	4	44.44%	5	55.56%	13	48.15%	14	51.85%	10	50.00%	10	50.00%
never	29	30.21%	67	69.79%	80	44.20%	101	55.80%	22	29.33%	53	70.67%
recently	8	21.05%	30	78.95%	20	42.55%	27	57.45%	3	14.29%	18	85.71%
just	24	72.73%	9	27.27%	68	90.67%	7	9.33%	13	41.94%	18	58.06%
since	18	85.71%	3	14.29%	33	67.35%	16	32.65%	16	66.67%	8	33.33%

Appendix D. Cluster dendrograms with selected variables excluded: ICE-HK

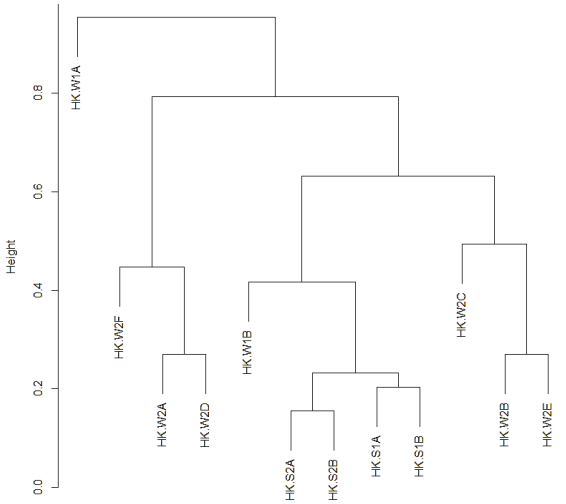


Figure 8.a. Cluster dendrogram of similarity across ICE text categories in HKE: without temporal adverbial specification frequency

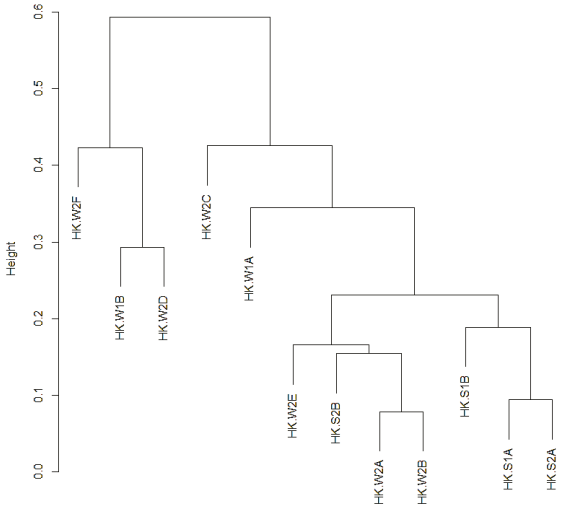


Figure 8.b. Cluster dendrogram of similarity across ICE text categories in HKE: without temporal adverbial categories

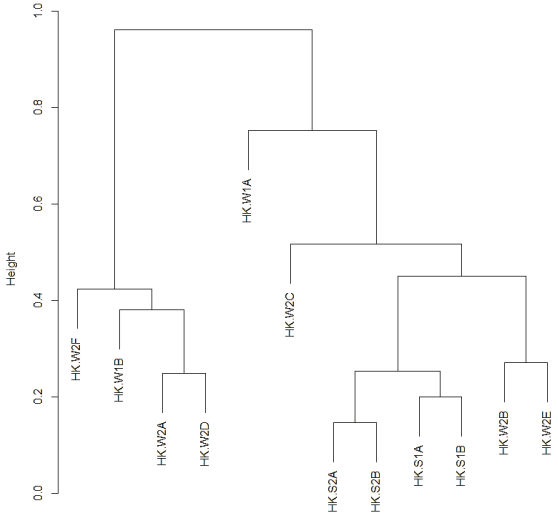


Figure 8.c. Cluster dendrogram of similarity across ICE text categories in HKE: without sentence type

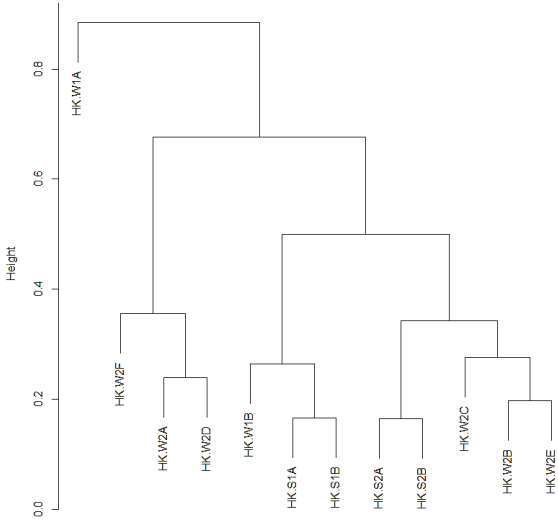


Figure 8.d. Cluster dendrogram of similarity across ICE text categories in HKE: without semantics

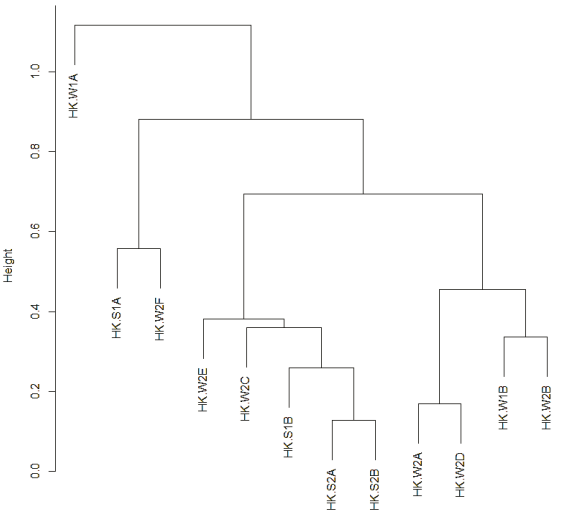


Figure 8.e. Cluster dendrogram of similarity across ICE text categories in HKE: without preceding TR form

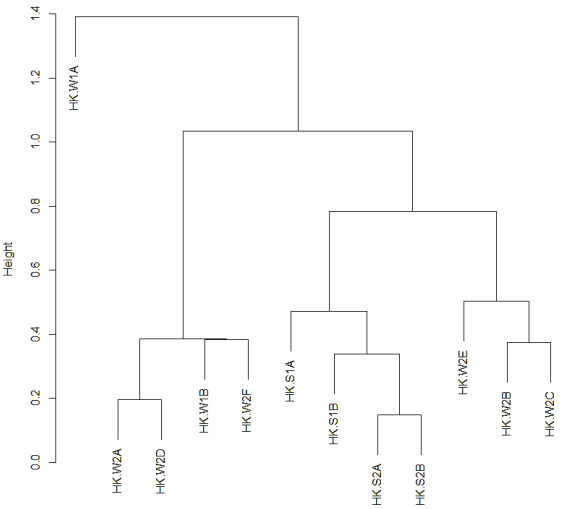


Figure 8.f. Cluster dendrogram of similarity across ICE text categories in HKE: without sentence type and preceding TR form

References

Agnihotri, Rama K. 1988. *Tense in Indian English*. New Delhi: Bahri.

Ahlqvist, Anders. 2000. Preterite and perfect in modern English. In Olga M. Tomić & Milorad Radovanović (eds.), *History and perspectives of language study*, 171–181. Amsterdam: Benjamins.

Åkerlund, Alfred. 1911. *On the history of the definite tenses in English*. Lund: Berlingska.

Algeo, John. 2006. *British or American English? A handbook of word and grammar patterns*. Cambridge: Cambridge University Press.

Allerton, David J., Paul Skandera & Cornelia Tschichold. 2002. Introduction. In David J. Allerton, Paul Skandera & Cornelia Tschichold (eds.), *Perspectives on English as a world language*, XI–XIV. Basel: Schwabe.

Allerton, David J. 2008. Swiss English, German English and American English: In grammatical alliance against traditional British English? In Miriam A. Locher & Jürg Strässler (eds.), *Standards and norms in the English language*, 24–45. Berlin: Mouton de Gruyter.

Ammon, Ulrich. 2005. Pluricentric and divided languages. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An international handbook of the science of language and society*, 1536–1543. Berlin: Mouton de Gruyter.

Anderson, Lloyd B. 1982. The 'Perfect' as a universal and as a language-particular category. In Paul J. Hopper (ed.), *Tense-aspect: Between semantics & pragmatics*, 227–264. Amsterdam: Benjamins.

Anderwald, Lieselotte. 2009. *The morphology of English dialects: Verb formation in non-standard English*. Cambridge: Cambridge University Press.

Ansaldo, Umberto. 2004. The evolution of Singapore English: Finding the matrix. In Lisa Lim (ed.), *Singapore English: A grammatical description*, 127–149. Amsterdam: Benjamins.

Ansaldo, Umberto. 2010. Contact and Asian varieties of English. In Raymond Hickey (ed.), *The handbook of language contact*, 498–517. Chichester: Wiley-Blackwell.

Appaduraj, Arjun. 1996. *Modernity at large: Cultural dimensions of globalization*. Minneapolis: University of Minnesota Press.

Askedal, John Ole. 2010. Periphrastische Passiv- und Perfektkonstruktionen in den germanischen Sprachen: Unterschiedliche Grammatikalisierungsstufen und ihre areale Distribution. In Antje Dammel, Sebastian Kürschner & Damaris Nübling (eds.), *Kontrastive Germanistische Linguistik. Vol. 2*, 553–583. Hildesheim: Olms.

Bache, Carl. 1994. Verbal categories, form-meaning relationships and the English perfect. In Carl Bache, Hans Basboell & Carl-Erik Lindberg (eds.), *Tense, aspect and action: Empirical and theoretical contributions to language typology*, 43–60. Berlin: Mouton de Gruyter.

Balasubramanian, Chandrika. 2009. *Register variation in Indian English*. Amsterdam: Benjamins.

Bao, Zhiming. 2005. The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics* 41(2). 237–267.

Bardovi-Harlig, Kathleen. 1997. Another piece of the puzzle: The emergence of the present perfect. *Language Learning* 47. 375–422.

Bardovi-Harlig, Kathleen. 2000. *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bartsch, Renate. 1995. *Situations, tense and aspect*. Berlin: Mouton de Gruyter.

Bauer, Gero. 1970. The English 'Perfect' reconsidered. *Journal of Linguistics* 6. 189–198.

Baugh, Albert C. & Thomas Cable. 2013. *A history of the English language*. London: Routledge.

Bautista, Maria L. S. 2008. Investigating the grammatical features of Philippine English. In Maria L. S. Bautista & Kingsley Bolton (eds.), *Philippine English: Linguistic and literary perspectives*, 201–218. Hong Kong: Hong Kong University Press.

Bhatt, Rakesh M. 2004. Indian English: Syntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 1016–1030. Berlin: Mouton de Gruyter.

Bhatt, Rakesh M. 2010. World Englishes, globalization and the politics of conformity. In Mukul Saxena & Tope Omoniyi (eds.), *Contending with globalization in World Englishes*, 93–112. Bristol: Multilingual Matters.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1990. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing* 5(4). 257–269.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge: Cambridge University Press.

Biewer, Carolin. 2008. South Pacific Englishes: Unity and diversity in the usage of the present perfect. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen (eds.), *The dynamics of linguistic variation*, 203–219. Amsterdam: Benjamins.

Binnick, Robert I. 1991. *Time and the verb: A guide to tense and aspect*. Oxford: Oxford University Press.

Binnick, Robert I. 2009. Tense and aspect. In Frank Brisard, Jef Verschueren & Jan-Ola Östman (eds.), *Grammar, meaning and pragmatics*, 268–288. Amsterdam: Benjamins.

Bolton, Kingsley. 2005. Hong Kong English: Autonomy and creativity. In Kingsley Bolton (ed.), *Hong Kong English: Autonomy and creativity*, 1–25. Hong Kong: Hong Kong University Press.

Bongartz, Christiane M. & Sarah Buschfeld. 2011. English in Cyprus: Second language variety or learner English? In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 35–54. Amsterdam: Benjamins.

Boogart, Ronny. 1999. *Aspect and temporal ordering: A contrastive analysis of Dutch and English*. Amsterdam: Vrije Universiteit PhD.

Bowie, Jill & Bas Aarts. 2012. Change in the English infinitival perfect construction. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 200–210. Oxford: Oxford University Press.

Bowie, Jill, Sean Wallis & Bas Aarts. 2013. The perfect in spoken British English. In Bas Aarts, Joan Close, Geoffrey N. Leech & Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 318–350. Cambridge: Cambridge University Press.

Brinton, Laurel J. 1988. *The development of English aspectual systems: Aspectualizers and post-verbal particles*. Cambridge: Cambridge University Press.
Brinton, Laurel J. 1999. Rev. of The Perfect and the Preterite in Contemporary and Earlier English, by Johan Elsness. *English Language and Linguistics* 3(2). 357–361.
Bruthiaux, Paul. 2003. Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics* 13(2). 159–178.
Brutt-Griffler, Janina. 2002. *World English*. Bristol: Multilingual Matters.
Bybee, Joan. 2011. Markedness: Iconicity, economy, and frequency. In Jae Jung Son (ed.), *The Oxford handbook of linguistic typology*, 131–147. Oxford: Oxford University Press.
Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13(1). 51–103.
Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
Canavan, John R. 1983. *The English tense system*. Bonn: Bouvier.
Chambers, Jack C. 2003. *Sociolinguistic theory: Linguistic variation and its social implications*. Malden: Blackwell.
Clyne, Michael G. 2005. Pluricentric language. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An international handbook of the science of language and society*, 296–300. Berlin: Mouton de Gruyter.
Collins, Peter & Pam Peters. 2004. Australian English: Morphology and syntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 593–610. Berlin: Mouton de Gruyter.
Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.
Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
Conrad, Andrew W. 1996. The international role of English: The state of the discussion. In Joshua A. Fishman (ed.), *Post-imperial English: Status change in former British and American colonies, 1940–1990*, 13–36. Berlin: Mouton de Gruyter.

Corrigan, Karen P. 2010. *Irish English*. Edinburgh: Edinburgh University Press.
Croft, William. 2012. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press.
Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.
Dahl, Östen. 1995. Areal tendencies in tense-aspect systems. In Pier M. Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality*: *Vol. 2. Typological perspectives*, 11–27. Torino: Rosenberg & Sellier.
Dahl, Östen & Eva Hedin. 2000. Current relevance and event reference. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 385–401. Berlin: Mouton de Gruyter.
Dahl, Östen & Viveka Velupillai. 2011a. Tense and aspect. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, supplement 7. <http://wals.info/supplement/7> (08 December 2011).
Dahl, Östen & Viveka Velupillai. 2011b. Perfective/imperfective aspect. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, chapter 65. <http://wals.info/chapter/65> (08 December 2011).
Dahl, Östen & Viveka Velupillai. 2011c. The perfect. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, chapter 68. <http://wals.info/chapter/68> (08 December 2011).
Dahl, Östen & Viveka Velupillai. 2011d. The past tense. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library, chapter 66. <http://wals.info/chapter/66> (08 December 2011).
Davies, Mark. 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25(4). 1–18.
Davies, Mark. 2013. Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries. <corpus2.byu.edu/glowbe> (18 February 2014).
Davydova, Julia. 2008. *Preterite and present perfect in Irish English*: *Determinants of variation*. Hamburg: SFB 538.

Davydova, Julia. 2011. *The present perfect in non-native Englishes: A corpus-based study of variation*. Berlin: Mouton de Gruyter.

Davydova, Julia. 2013. Detecting historical continuity in a linguistically diverse urban area. The present perfect in modern Singapore English. In Joana Duarte & Ingrid Gogolin (eds.): *Linguistic superdiversity in urban areas: Research approaches*, 193–225. Amsterdam: Benjamins.

Davydova, Julia, Michaela Hilbert, Lukas Pietsch & Peter Siemund. 2011. Comparing varieties of English: Problems and perspectives. In Peter Siemund (ed.), *Linguistic universals and language variation*, 291–323. Berlin: Mouton de Gruyter.

Declerck, Renaat. 1991. *Tense in English*. London: Routledge.

Declerck, Renaat. 2006. *The grammar of the English verb phrase*: *Vol I. The Grammar of the English tense system. A comprehensive analysis*. Berlin: Mouton de Gruyter.

Depraetere, Ilse. 1996. *The tense system in English relative clauses*: *A corpus-based analysis*. Berlin: Mouton de Gruyter.

Depraetere, Ilse. 1998. On the resultative character of present perfect sentences. *Journal of Pragmatics* 29. 597–613.

Deterding, David. 2008. *Singapore English*. Edinburgh: University of Edinburgh Press.

Deuber, Dagmar. 2009. 'The English we speaking': Morphological and syntactic variation in educated Jamaican speech. *Journal of Pidgin and Creole Languages* 24(1). 1–52.

Diessel, Holger. 2011. Grammaticalization and language acquisition. In Bernd Heine & Heiko Narrog (eds.), *Handbook of grammaticalization*, 130–141. Oxford: Oxford University Press.

Dinsmore, John. 1981. Tense choice and time specification in English. *Linguistics* 19. 475–494.

Dinsmore, John. 1991. The logic and functions of the English past and perfect. In Carol P. Georgopoulos & Roberta Ishihara (eds.), *Interdisciplinary approaches to language*, 101–117. Dordrecht: Kluwer.

Drinka, Bridget. 2003. Areal factors in the development of the European periphrastic perfect. *Word* 54(1). 1–38.

Dryer, Matthew S. & Martin Haspelmath, eds. 2011. *The world atlas of language structures online*. Munich: Max Planck Digital Library.

Dua, Hans R. 1996. The spread of English in India: Politics of language conflict and language power. In Joshua A. Fishman (ed.), *Post-imperial English: Status change in former British and American colonies, 1940–1990*, 557–588. Berlin: Mouton de Gruyter.

Dubois, Betty L. 1972. *The meanings and the distribution of the perfect in present-day American English writing*. Albuquerque: University of New Mexico PhD.

Durrleman-Tame, Stephanie. 2008. *The syntax of Jamaican Creole: A cartographic perspective*. Amsterdam: Benjamins.

Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27(2). 315–330.

Edwards, Alison. 2010. Dutch English: tolerable, taboo, or about time too? *English Today* 26(1). 19–25.

Elliott, Anthony. 2010. *Globalization: A reader*. New York: Routledge.

Elsness, Johan. 1997. *The perfect and the preterite in contemporary and earlier English*. Berlin: Mouton de Gruyter.

Elsness, Johan. 2009a. The perfect and the preterite in Australian and New Zealand English. In Pam Peters, Peter Collins & Adam Smith (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and beyond*, 89–114. Amsterdam: Benjamins.

Elsness, Johan. 2009b. The present perfect and the preterite. In Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English*, 228–245. Cambridge: Cambridge University Press.

Elsness, Johan. 2012. The present perfect in late Modern English: A longitudinal look. Talk given at ICAME 33: Corpora at the centre and crossroads of English linguistics. Leuven, May 2012.

Engel, Dulcie M. & Marie-Eve A. Ritz. 2000. The use of the present perfect in Australian English. *Australian Journal of Linguistics* 20(2). 119–140.

Evans, Stephen. 2009. The evolution of the English-language speech community in Hong Kong. *English World-Wide* 30(3). 278–301.

eWAVE Feature 88. 2012. Wider range of uses of progressive be + V-ing than in StE: extension to stative verbs. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*.

Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/88> (08 November 2012).
eWAVE Feature 89. 2012. Wider range of uses of progressive be + V-ing than in StE: extension to habitual contexts. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/89> (08 November 2012).
eWAVE Feature 97. 2012. Medial object perfect. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/97> (05 September 2012).
eWAVE Feature 98. 2012. After-perfect. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/98> (10 May 2012).
eWAVE Feature 100. 2012. Levelling of the difference between present perfect and simple past: present perfect for StE simple past. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/100> (05 September 2012).
eWAVE Feature 111. 2012. Past tense/anterior marker been. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/111> (05 January 2013).
eWAVE Feature 131. 2012. Levelling of past tense/past participle verb forms: past participle replacing the past tense form. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/parameters/131> (08 February 2013).
Fenn, Peter. 1987. *A semantic and pragmatic examination of the English perfect*. Tübingen: Narr.
Fishman, Joshua A. 1996a. Introduction: Some empirical and theoretical issues. In Joshua A. Fishman (ed.), *Post-imperial English*: *Status change*

in former British and American colonies, 1940–1990, 3–12. Berlin: Mouton de Gruyter.
Fishman, Joshua A. 1996b. Summary and interpretation: Post-imperial English 1940–1990. In Joshua A. Fishman (ed.), *Post-imperial English: Status change in former British and American colonies, 1940–1990*, 623–641. Berlin: Mouton de Gruyter.
Filppula, Markku. 1997. The Influence of Irish on perfect marking in Hiberno-English: The case of the 'extended-now' perfect. In Jeffrey L. Kallen (ed.), *Focus on Ireland*, 51–71. Amsterdam: Benjamins.
Filppula, Markku. 2004. Irish English: Morphology and syntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 73–101. Berlin: Mouton de Gruyter.
Filppula, Markku. 2012. Irish English. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/languages/3> (20 June 2012).
Fritz, Clemens. 2006. Resilient or yielding? Features of Irish English syntax and aspect in early Australia. In Terttu Nevalainen, Juhani Klemola & Mikko Laitinen (eds.), *Types of variation*, 281–301. Amsterdam: Benjamins.
Garcia, Erica C. 1985. *History of the English tense system*. Ann Arbor: University Microfilms.
Giorgi, Alessandra & Fabio Pianesi. 1997. *Tense and Aspect: From semantics to morphosyntax*. Oxford: Oxford University Press.
Gonzales, Andrew, Thelma V. Jambalos & Maria Corona Salcedo Romero. 2003. *Three studies on Philippine English across generations*. Manila: Linguistic Society of the Philippines.
Görlach, Manfred. 1990. *Studies in the history of the English language*. Heidelberg: Winter.
Görlach, Manfred. 1995. *More Englishes*. Amsterdam: Benjamins.
Görlach, Manfred. 2002. *Still more Englishes*. Amsterdam: Benjamins.
Graddol, David. 2006. *English next*. London: British Council.
Granger, Sylviane, Estelle Dagneaux, Fanny Meunier & Magali Paquot. 2009. *International corpus of learner English v2*. Louvain-la-Neuve: Presses universitaires de Louvain.

Greenbaum, Sidney, ed. 1996a. *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Greenbaum, Sidney. 1996b. Introducing ICE. In Sidney Greenbaum (ed.) *Comparing English worldwide: The International Corpus of English*, 3–12. Oxford: Clarendon Press.

Gries, Stefan Thomas. 2006. Exploring variability within and between corpora: Some methodological considerations. *Corpora* 1(2). 109–151.

Gries, Stefan Thomas & Anatol Stefanowitsch. 2010. Cluster analysis and the identification of collexeme classes. In John Newman & Sally Rice (eds.), *Empirical and experimental methods in cognitive/functional research*, 73–90. Stanford: CSLI.

Guenthner, Franz. 1977. Remarks on the present perfect in English. In Christian Rohrer (ed.), *On the logical analysis of tense and aspect*, 83–97. Tübingen: Narr.

Gut, Ulrike. 2009. Past tense marking in Singapore English verbs. *English World-Wide* 30(3). 262–277.

Hackert, Stephanie. 2008. Counting and coding the past: Circumscribing the variable context in quantitative analyses of past inflection. *Language Variation and Change* 20. 127–153.

Halford, Brigitte. 2002. Canadian English: Distinct in North America? In David J. Allerton, Paul Skandera & Cornelia Tschichold (eds.), *Perspectives on English as a world language*, 31–44. Basel: Schwabe.

Halliday, M.A.K & Christian Matthiessen. 2004. *An introduction to functional grammar*. London: Hodder.

Hantson, André. 2005. The English perfect and the anti-perfect *used to* viewed from a comparative perspective. *English Studies* 86(3). 245–268.

Harder, Peter. 1997. *Functional semantics*: *A theory of meaning, structure and tense in English*. Berlin: Mouton de Gruyter.

Harris, John. 1984. Syntactic variation and dialect divergence. *Journal of Linguistics* 20(2). 303–327.

Hartford, Beverly A. S. 1993. Tense and aspect in the news discourse of Nepali English. *World Englishes* 12(1). 1–13.

Hay, Jennifer, Margaret Maclagan & Elizabeth Gordon. 2008. *New Zealand English*. Edinburgh: Edinburgh University Press.

Hickey, Raymond. 2004. Introduction. In Raymond Hickey (ed.), *Legacies of colonial English: Studies in transported dialects*, 1–30. Cambridge: Cambridge University Press.

Hickey, Raymond. 2010a. Contact and language shift. In Raymond Hickey (ed.), *The handbook of language contact*, 151–169. Chichester: Wiley-Blackwell.

Hickey, Raymond. 2010b. Language contact: Reconsideration and reassessment. In Raymond Hickey (ed.), *The handbook of language contact*, 1–28. Chichester: Wiley-Blackwell.

Hickey, Raymond, ed. 2012. *Standards of English. Codified varieties around the world*. Cambridge: Cambridge University Press.

Hoffmann, Sebastian, Andrea Sand & Peter Tan. 2012. The Corpus of Historical Singapore English – A First Pilot Study on Data from the 1950s and 1960s. Talk given at ICAME 33: Corpora at the centre and crossroads of English linguistics. Leuven, May 2012.

Huddleston, Rodney. 1996. The English perfect as a secondary past tense. In Bas Aarts & Charles F. Meyer (eds.), *The verb in contemporary English: Theory and description*, 102–122. Cambridge: Cambridge University Press.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hundt, Marianne. 1998. *New Zealand English grammar, fact or fiction?* Amsterdam: Benjamins.

Hundt, Marianne. 2002. English in New Zealand. In David J. Allerton, Paul Skandera & Cornelia Tschichold (eds.), *Perspectives on English as a world language*, 63–78. Basel: Schwabe.

Hundt, Marianne. 2009. Global feature – local norms? A case study on the progressive passive. In Thomas Hoffmann & Lucia Siebers (eds.), *World Englishes – problems, properties and prospects*, 287–308. Amsterdam: Benjamins.

Hundt, Marianne. 2013. Diversification of English: Old, new and emerging epicentres. In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language*, 182–203. Cambridge: Cambridge University Press.

Hundt, Marianne & Christian Mair. 1999. 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2). 221–242.
Hundt, Marianne & Carolin Biewer. 2007. The dynamics of inner and outer circle varieties in the South Pacific and Asia. In Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.), *Corpus linguistics and the web*, 249–269. Amsterdam: Rodopi.
Hundt, Marianne & Katrin Vogel. 2011. Overuse of the progressive in ESL and learner Englishes – fact or fiction? In Joybrato Mukherjee & Marianne Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 145–165. Amsterdam: Benjamins.
Hundt, Marianne & Nicolas Smith. 2009. The present perfect in British and American English: Has there been a change, recently? *ICAME Journal* 33. 45–63.
Hundt, Marianne, Jennifer Hay & Elizabeth Gordon. 2004. New Zealand English: Morphosyntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 560–592. Berlin: Mouton de Gruyter.
Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267.
Iatridou, Sabine. 2003. A little bit more on the English perfect. In Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.), *Perfect explorations*, 133–152. Berlin: Mouton de Gruyter.
Inoue, Kyoko. 1979. An analysis of the English present perfect. *Linguistics* 17. 561–589.
Jaszczolt, Kasia M. 2009. *Representing time: An essay on temporality as modality*. Oxford: Oxford University Press.
Jenkins, Jennifer. 2009. *World Englishes*. London: Routledge.
Jespersen, Otto. 1931. *A modern English grammar on historical principles: Part iv. Syntax. Third volume. Time and tense*. Heidelberg: Winter.
Joos, Martin. 1964. *The English verb*. Madison: University of Wisconsin Press.
Kachru, Braj B. 1988. The sacred cows of English. *English Today* 16. 3–8.

Kachru, Yamuna & Cecil L. Nelson. 2006. *World Englishes in Asian contexts*. Hong Kong: Hong Kong University Press.

Kachru, Yamuna & Larry E. Smith. 2008. *Cultures, contexts, and World Englishes*. New York: Routledge.

Kallen, Jeffrey L. 1989. Tense and aspect categories in Irish English. *English World-Wide* 10(1). 1–39.

Katz, Graham. 2003. On the stativity of the English perfect. In Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.), *Perfect explorations*, 205–234. Berlin: Mouton de Gruyter.

Kirk, John M. & Jeffrey L. Kallen. 2006. Irish Standard English: How Celticised? How standardised? In Hildegard L. C. Tristram (ed.), *The interface between English and the Celtic languages: Proceedings of the fourth International Colloquium on the "Celtic Englishes" held at the University of Potsdam in Golm (Germany) from 22–26 September 2004*, 88–113. Potsdam: Universitäts-Verlag Potsdam.

Kirk, John M. & Jeffrey L. Kallen. 2007. Assessing celticity in a corpus of standard Irish English. In Hildegard L. C. Tristram (ed.), *The Celtic languages in contact: Papers from the workshop within the framework of the XIII International Congress of Celtic Studies, Bonn, 26–27 July 2007*, 270–298. Potsdam: Universitäts-Verlag Potsdam.

Kirsten, Hans. 1994. *Englische Verbformen: Bedeutung und kommunikative Leistung*. Essen: Die Blaue Eule.

Klein, Wolfgang. 1992. The present perfect puzzle. *Language* 68(3). 525–552.

Klein, Wolfgang. 1994. *Time in language*. London: Routledge.

Klein, Wolfgang. 2009a. Concepts of time. In Wolfgang Klein & Ping Li (eds.), *The expression of time*, 5–38. Berlin: Mouton de Gruyter.

Klein, Wolfgang. 2009b. How time is encoded. In Wolfgang Klein & Ping Li (eds.), *The expression of time*, 39–81. Berlin: Mouton de Gruyter.

Klein, Wolfgang. 2010. On times and arguments. *Linguistics* 48(6). 1221–1253.

Klein, Wolfgang & Heinz Vater. 1998. The perfect in English and German. In Leonid I. Kulikov & Heinz Vater (eds.), *Typology of verbal categories: Papers presented to Vladimir Nedjalkov on the occasion of his 70th birthday*, 215–235. Tübingen: Niemeyer.

König, Ekkehard. 1995. On analyzing the tense-aspect system of English: A state-of-the-art report. In Wolfgang Riehle & Hugo Keiper (eds.), *Anglistentag 1994 Graz*: *Proceedings*, 153–169. Tübingen: Niemeyer.

Köppl, Sebastian. 1983a. Englisch in Australien und Neuseeland. In Sebastian Köppl, Josef Schmied, Edgar W. Schneider & Wolfgang Viereck (eds.), *Englisch, Formen und Funktionen einer Weltsprache*: *Ausstellung des Lehrstuhls für Englische Sprachwissenschaft und Mediävistik und der Universitätsbibliothek, Univ. Bamberg, 19. Mai - 30. Juni 1983*, 145–154. Bamberg: Lehrstuhl für Engl. Sprachwiss. u. Mediävistik.

Köppl, Sebastian. 1983b. Englisch in Süd- und Südostasien. In Sebastian Köppl, Josef Schmied, Edgar W. Schneider & Wolfgang Viereck (eds.), *Englisch, Formen und Funktionen einer Weltsprache*: *Ausstellung des Lehrstuhls für Englische Sprachwissenschaft und Mediävistik und der Universitätsbibliothek, Univ. Bamberg, 19. Mai - 30. Juni 1983*, 117–143. Bamberg: Lehrstuhl für Engl. Sprachwiss. u. Mediävistik.

Kortmann, Bernd. 1991. The triad 'tense-aspect-Aktionsart'. *Belgian Journal of Linguistics* 6. 9–30.

Kortmann, Bernd. 1995. Compositionality and the perfect. In Wolfgang Riehle & Hugo Keiper (eds.), *Anglistentag 1994 Graz*: *Proceedings*, 183–199. Tübingen: Niemeyer.

Kortmann, Bernd, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton, eds. 2004. *A handbook of varieties of English*. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Edgar W. Schneider. 2006. *Varieties of English: A multimedia reference tool*. Berlin: Mouton de Gruyter. <http://www.varieties.mouton-content.com/> (23 January 2013).

Kortmann, Bernd & Benedikt Szmrecsanyi. 2004. Global synopsis: Morphological and syntactic variation in English. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 1142–1202. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Benedikt Szmrecsanyi. 2009. World Englishes between simplification and complexification. In Thomas Hoffmann & Lucia Siebers (eds.), *World Englishes – problems, properties and prospects*, 265–285. Amsterdam: Benjamins.

Kortmann, Bernd & Benedikt Szmrecsanyi. 2011. Parameters of morphosyntactic variation in World Englishes: Prospects and limitations of searching for universals. In Peter Siemund (ed.), *Linguistic universals and language variation*, 264–290. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Christoph Wolk. 2012. Morphosyntactic variation in the anglophone world: A global perspective. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 906–936. Berlin: Mouton de Gruyter.

Kortmann, Bernd & Kerstin Lunkenheimer, eds. 2011. *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Koziol, Herbert. 1958. Zum Gebrauch des Present Perfect und des Past Tense. *Die Neueren Sprachen* 7. 497–506.

Kretzschmar, William A. Jr. 2012. Complex systems and sociolinguistics. Talk given at the University of Cambridge Linguistics Society. Cambridge, March 2012.

Krug, Manfred. 2011. Auxiliaries and grammaticalization. In: Bernd Heine & Heiko Narrog (eds.), *Handbook of grammaticalization*, 547–558. Oxford: Oxford University Press.

Krug, Manfred & Anna Rosen. 2012. Standards of English in Malta and the Channel Islands. In: Raymond Hickey (ed.), *Standards of English: Codified varieties around the world*, 117–138. Cambridge: Cambridge University Press.

Krug, Manfred, Lisa Bonnici & Michaela Hilbert. 2012. Maltese English. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 653–668. Berlin: Mouton de Gruyter.

Krug, Manfred, Julia Schlüter & Anette Rosenbach. 2013. Introduction: Investigating language variation and change. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 1–14. Cambridge: Cambridge University Press.

Krug, Manfred & Katrin Sell. 2013. Designing and Conducting Interviews and Questionnaires. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 69–98. Cambridge: Cambridge University Press.

Kuhn, Steven T. 1989. Tense and time. In Dov M. Gabbay & Franz Guenthner (eds.), *Handbook of philosophical logic*: *Volume IV*. *Topics in the philosophy of language*, 513–552. Dordrecht: Reidel.

Labov, William. 1978. *The study of NonStandard English*. Urbana: National Council of Teachers of English.

Labov, William. 1990. On the adequacy of natural languages: The development of tense. In John V. Singler (ed.), *Pidgin and creole tense-mood-aspect systems*, 1–58. Amsterdam: Benjamins.

Landman, Fred. 2008. 1066: On the differences between the tense-perspective-aspect systems of English and Dutch. In Susan Rothstein (ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect*, 107–166. Amsterdam: Benjamins.

Lange, Claudia. 2012. *The syntax of spoken Indian English*. Amsterdam: Benjamins.

Larntz, Kinley. 1978. Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics. *Journal of the American Statistical Association* 73(362). 253–263.

Leech, Geoffrey N. 2004. *Meaning and the English verb*. Harlow: Pearson Longman.

Leech, Geoffrey N., Marianne Hundt, Christian Mair & Nicolas Smith. 2009. *Change in contemporary English*: *A grammatical study*. Cambridge: Cambridge University Press.

Leitner, Gerhard. 1992. English as a pluricentric language. In Michael Clyne (ed.), *Pluricentric languages: Differing norms in different nations*, 179–237. Berlin: Mouton de Gruyter.

Lian, Ho M. 2003. Past tense marking in Singapore English. In David Deterding, Ee Ling Low & Adam Brown (eds.), *English in Singapore*: *Research on grammar*, 39–47. Singapore: McGraw Hill.

Lindstedt, Juoko. 2000. The perfect – aspectual, temporal and evidential. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 365–383. Berlin: Mouton de Gruyter.

Liszka, Sarah A. 2003. *The development of tense and the present perfect in second language English*. Colchester: University of Essex PhD.

Lucko, Peter. 1981. *Das Perfekt im Deutschen und das Present Perfect im Englischen – eine konfrontative Analyse*. Berlin: Humboldt Universität PhD.

Lunkenheimer, Kerstin. 2012. Tense and aspect. In Raymond Hickey (ed.), *Areal features of the Anglophone world*, 329–353. Berlin: Mouton de Gruyter.

Mair. Christian. 2013. The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars. *English World-Wide* 34(3). 253–278.

Manarpaac, Danilo. 2003. 'When I was a child I spake as a child': Reflecting on the limits of a nationalist language policy. In Christian Mair (ed.), *The politics of English as a world language*, 479–492. Amsterdam: Rodopi.

Manning, Christopher D. & Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. London: MIT Press.

Markus, Manfred. 1977. *Tempus und Aspekt: Zur Funktion von Präsens, Präteritum und Perfekt im Englischen und Deutschen*. München: Fink.

Matthews, Richard. 1987. Present perfect tenses: Towards an integrated functional account. In Alfred Schopf (ed.), *Reference time, tense and adverbs*, 111–176. Berlin: Mouton de Gruyter.

Maule, David. 1991. *The naked verb: The meaning of the English verb tenses*. London: Macmillan.

McArthur, Tom. 1987. The English languages? *English Today* 11. 9–11.

McArthur, Tom. 1998. *The English languages*. Cambridge: Cambridge University Press.

McArthur, Tom. 2002. *The Oxford guide to World English*. Oxford: Oxford University Press.

McCafferty, Kevin. 2005. Future, perfect – and past? Changing uses of *be after V-ing* in Irish English. In Gulbrand Alhaug, Endre Mørck & Aud-Kristi Pedersen (eds.), *Mot rikare mål å trå: Festskrift til Trove Bull*, 241–251. Oslo: Novus.

McCawley, James D. 1983. Tense and time reference in English. In Charles J. Fillmore & D. Terence Langendoen (eds.), *Studies in linguistic semantics*, 96–113. New York: Holt, Rinehart and Winston.

McCoard, Robert W. 1978. *The English perfect: Tense-choice and pragmatic inferences*. Amsterdam: North-Holland.

McGilvray, James A. 1991. *Tense, reference, and worldmaking*. Montreal: McGill-Queen's University Press.

McFadden, Thomas & Artemis Alexiadou. 2010. Perfects, resultatives, and auxiliaries in earlier English. *Linguistic Enquiry* 41(3). 389–425.
McMahon, April & Robert McMahon. 2005. *Language classification by numbers*. Oxford: Oxford University Press.
McMahon, April & Warren Maguire. 2013. Computing linguistic distances between varieties. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 421–432. Cambridge: Cambridge University Press.
Mesthrie, Rajend & Rakesh M. Bhatt. 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.
Meyer, Matthias. 1992. *Das englische Perfekt*. Tübingen: Niemeyer.
Meyer, Matthias. 1995. Past tense and present perfect in the Brown and the LOB corpora. In Wolfgang Riehle & Hugo Keiper (eds.), *Anglistentag 1994 Graz: Proceedings*, 201–228. Tübingen: Niemeyer.
Michaelis, Laura A. 1994. The ambiguity of the English present perfect. *Journal of Linguistics* 30(1). 111–157.
Michaelis, Laura A. 1998. *Aspectual grammar and past time reference*. London: Routledge.
Michaelis, Laura A. 2006. Time and tense. In Bas Aarts (ed.), *The handbook of English linguistics*, 220–243. Malden: Blackwell.
Miller, Jim. 2000. The Perfect in spoken and written English. *Transactions of the Philological Society* 98(2). 323–352.
Miller, Jim. 2004a. Perfect and resultative constructions in spoken and non-standard English. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and down the cline: The nature of grammaticalization*, 229–246. Amsterdam: Benjamins.
Miller, Jim. 2004b. Problems for typology: Perfects and resultatives in spoken and non-standard English and Russian. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 305–334. Berlin: Mouton de Gruyter.
Mindt, Dieter. 2000. *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
Minnich, Anja. 2004. *Conceptualising English: A metaphorology of models in the discipline of World Englishes*. Cambridge: University of Cambridge PhD.

Mittwoch, Anita. 2008. The English resultative perfect and its relationship to the experiential perfect and the simple past tense. *Linguistics and Philosophy* 31. 323–351.
Modiano, Marko. 1999. Standard English(es) and educational practices for the world's lingua franca. *English Today* 15(4). 3–13.
Moens, Marc. 1987. *Tense, aspect and temporal reference*. Edinburgh: University of Edinburgh PhD.
Moens, Marc & Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14(2). 15–28.
Moisl, Hermann. 2010. Variable scaling in cluster analysis of linguistic data. *Corpus Linguistics and Linguistic Theory* 6(1). 75–103.
Molsing, Karina V. 2006. The present perfect at the semantics/pragmatics interface: American English and Brazilian Portuguese. In Christian Ebert & Cornelia Endriss (eds.), *Proceedings of the Sinn und Bedeutung 10: Annual meeting of the Gesellschaft für Semantik*, 239–250. Berlin: ZAS.
Mugler, Alfred. 1988. *Tempus und Aspekt als Zeitbeziehungen*. München: W. Fink.
Mukherjee, Joybrato. 2005. Review of: Norbert Schlüter. Present perfect: Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht. *ICAME Journal* 29. 198–203.
Mukherjee, Joybrato & Marianne Hundt, eds. 2011. *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*. Amsterdam: Benjamins.
Mukherjee, Joybrato & Marco Schilk. 2012. Exploring variation and change in New Englishes: Looking into the International Corpus of English (ICE) and beyond. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 189–199. Oxford: Oxford University Press.
Muysken, Pieter. 2010. Scenarios for language contact. In Raymond Hickey (ed.), *The handbook of language contact*, 265–281. Chichester: Wiley-Blackwell.
Naumann, Ralf. 1998. A dynamic approach to the present perfect in English. *Theoretical Linguistics* 24(2-3). 185–218.

Nelson, Gerald. 1996. The design of the corpus. In Sidney Greenbaum (ed.) *Comparing English worldwide: The International Corpus of English*, 27–35. Oxford: Clarendon Press.

Nichols, Johanna & Tandy Warnow. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2(5). 760–820.

Nishiyama, Atsuko & Jean-Pierre Koenig. 2010. What is a perfect state? *Language* 86(3). 611–646.

Nordlander, Johan. 1997. *Towards a semantics of linguistic time: Exploring some basic time concepts with special reference to English and Krio*. Uppsala: Swedish Science Press.

Nordlander, Johan. 1998. *The English present perfect: A discourse marker of intra-sentential cohesion*. <http://www.eng.umu.se/johan/pdf/In%20Progress/Perfecta.pdf> (05 March 2011).

O'Keeffe, Anne & Carolina P. Amador Moreno. 2009. The pragmatics of the be + after + V-ing construction of Irish English. *Intercultural Pragmatics* 6(4). 517–534.

Omoniyi, Tope & Mukul Saxena. 2010. Introduction. In Mukul Saxena & Tope Omoniyi (eds.), *Contending with globalization in World Englishes*, 1–22. Bristol: Multilingual Matters.

Ota, Akira. 1963. *Tense and aspect of present-day American-English*. Tokyo: Kenkyusha.

Panitz, Florian. 1998. *Die temporalen Elemente des Englischen und deren Zeitbezug in fiktionalen narrativen Texten*. Tübingen: Niemeyer.

Panzner, Antje. 1995. *Das Pre-present in Kollokation mit Temporalangaben im britischen und amerikanischen Englisch der Gegenwart*. Frankfurt am Main: Lang.

Patrick, Peter L. 1999. *Urban Jamaican Creole*. Amsterdam: Benjamins.

Patrick, Peter L. 2004. Jamaican creole: Morphology and syntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 407–438. Berlin: Mouton de Gruyter.

Payre-Ficout, Coralie, Catherine Brissaud & Jean-Pierre Chevrot. 2009. The acquisition of the simple past/present perfect by French secondary school learners: When the form/function relations diverge in L1 and L2. Paper presented at the International conference on first and second languages: Exploring the relationship in pedagogy-related con-

texts. <http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2010/08/Payre-Ficout-et-al.doc> (15 May 2010).
Pennycook, Alastair. 1994. *The cultural politics of English as an international language*. London: Longman.
Pennycook, Alastair. 2007. *Global Englishes and transcultural flows*. London: Routledge.
Petersen, Christian H. 2004. *Formal and functional perspectives on tense and related categories*. Bern: Lang.
Phillipson, Robert. 1992. *Linguistic imperialism*. Oxford: Oxford University Press.
Phillipson, Robert. 2009. *Linguistic imperialism continued*. London: Routledge.
Pienemann, Manfred & Jörg-U. Kessler. 2012. Processability theory. In Susan M. Gass & Alison Mackey (eds.), *The Routledge handbook of second language acquisition*, 228–248. London: Routledge.
Pietsch, Lukas. 2009. Hiberno-English medial-object perfects reconsidered: A case of contact-induced grammaticalisation. *Studies in Language* 33(3). 528–568.
Platt, John T. 1982. English in Singapore, Malaysia, and Hong Kong. In Richard W. Bailey & Manfred Görlach (eds.), *English as a world language*, 384–414. Ann Arbor: University of Michigan Press.
Platt, John T., Heidi Weber & Mian-Lian Ho. 1984. *The new Englishes*. London: Routledge.
Poplack, Shana & Sali Tagliamonte. 2001. *African American English in the diaspora*. Malden: Blackwell.
Portner, Paul. 2003. The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy* 26. 459–510.
Pulgram, Ernst. 1987. Latin-romance, English, German past tenses and aspects: Defects and repairs. *Folia Linguistica Historica* 7(2). 381–397.
Quinn, Heidi. 1999. Variation in New Zealand English syntax and morphology. In Allan Bell & Koenraad Kuiper (eds.), *New Zealand English*, 173–197. Amsterdam: Benjamins.
Quirk, Randolph, Sidney Greenbaum, Geoffrey N. Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.

Rainer, Eva M. 1989. *Das Perfekt im Spätmittel- und Frühneuenglischen: Eine Frequenz- und Funktionsanalyse anhand von Brieftexten*. Innsbruck: Universität Innsbruck.

Ramanathan, Vaidehi, Bonny Norton & Alastair Pennycook. 2010. Preface. In Mukul Saxena & Tope Omoniyi (eds.), *Contending with globalization in World Englishes*, xv–xvii. Bristol: Multilingual Matters.

Rastall, Paul. 1999. Observations on the present perfect in English. *World Englishes* 18(1). 79–83.

Rathert, Monika. 2004. *Textures of time*. Berlin: Akademie.

Reichenbach, Hans. 1947. *Elements of symbolic logic*. London: Macmillan.

Ritz, Marie-Eve A. 2010. The perfect crime? Illicit uses of the present perfect in Australian police media releases. *Journal of Pragmatics* 42. 3400–3417.

Ritz, Marie-Eve A. 2012. Perfect tense and aspect. In: Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 881–907. Oxford: Oxford University Press.

Ritz, Marie-Eve A. & Dulcie M. Engel. 2008. 'Vivid narrative use' and the meaning of the present perfect in spoken Australian English. *Linguistics* 46(1). 131–160.

Roberts, Ian. 2007. *Diachronic syntax*. Oxford: Oxford University Press.

Rödel, Michael. 2007. *Doppelte Perfektbildungen und die Organisation von Tempus im Deutschen*. Tübingen: Stauffenburg.

Rohdenburg, Günter & Julia Schlüter, eds. 2009. *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press.

Romesburg, H. Charles. 1984. *Cluster analysis for researchers*. Belmont: Lifetime Learning.

Rosen, Anna. 2014. *Grammatical variation and change in Jersey English*. Amsterdam: Benjamins.

Rothstein, Björn. 2007. A modified ExtendedNow for the present perfect. In Louis de Saussure, Jaques Moeschler & Genoveva Puskás (eds.), *Recent advances in the syntax and semantics of tense, aspect and modality*, 89–107. Berlin: Mouton de Gruyter.

Rothstein, Björn. 2008. *The perfect time span: On the present perfect in German, Swedish and English*. Amsterdam: Benjamins.

Rothstein, Björn. 2011. Why the present perfect differs cross-linguistically: Some new insights. In Jesse Mortelmans, Tanja Mortelmans & Walter De Mulder (eds.), *From now to eternity*, 123–137. Amsterdam: Rodopi.

Sailaja, Pingali. 2009. *Indian English*. Edinburgh: Edinburgh University Press.

Salkie, Raphael. 1989. Perfect and pluperfect: What is the relationship? *Journal of Linguistics* 25. 1–34.

Sand, Andrea. 1999. *Linguistic variation in Jamaica: A corpus-based study of radio and newspaper usage*. Tübingen: Narr.

Sand, Andrea. 2002. English in the Caribbean. In David J. Allerton, Paul Skandera & Cornelia Tschichold (eds.), *Perspectives on English as a world language*, 79–91. Basel: Schwabe.

Sand, Andrea. 2005. The effects of language contact on the morpho-syntax of English. In Lilo Moessner & Christa M. Schmidt (eds.), *Proceedings Anglistentag 2004 Aachen*, 449–460. Trier: WVT.

Sankoff, Gilian. 2001. Linguistic outcomes of language variation. In Jack Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *Handbook of sociolinguistics*, 638–668. London: Blackwell.

Saraceni, Mario. 2010. *The relocation of English: Shifting paradigms in a global era*. Houndmills: Palgrave Macmillan.

Schaden, Gerhard. 2009. Present perfects compete. *Linguistics and Philosophy* 32. 115–141.

Schlüter, Norbert. 2000. The present perfect in British and American English: Selected results of an empirical analysis. In Christian Mair & Marianne Hundt (eds.), *Corpus linguistics and linguistic theory*, 313–320. Amsterdam: Rodopi.

Schlüter, Norbert. 2002a. *Present Perfect: Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Narr.

Schlüter, Norbert. 2002b. Temporal specification of the present perfect: A corpus-based study. In Pam Peters (ed.), *New frontiers of corpus research: Papers from the twenty first international conference on English language research on computerized corpora Sydney 2000*, 307–315. Amsterdam: Rodopi.

Schlüter, Norbert. 2006. How reliable are the results? Comparing corpus-based studies of the present perfect. *Zeitschrift für Anglistik und Amerikanistik* 54(2). 135–148.

Schmied, Josef. 1996. Second-language corpora. In Sidney Greenbaum (ed.), *Comparing English worldwide: The International Corpus of English*, 182–196. Oxford: Clarendon Press.

Schneider, Edgar W. 1983a. Englisch in der Karibik. In Sebastian Köppl, Josef Schmied, Edgar W. Schneider & Wolfgang Viereck (eds.), *Englisch, Formen und Funktionen einer Weltsprache*: *Ausstellung des Lehrstuhls für Englische Sprachwissenschaft und Mediävistik und der Universitätsbibliothek, Univ. Bamberg, 19. Mai - 30. Juni 1983*, 73–85. Bamberg: Lehrstuhl für Engl. Sprachwiss. u. Mediävistik.

Schneider, Edgar W. 1983b. Englisch in Nordamerika. In Sebastian Köppl, Josef Schmied, Edgar W. Schneider & Wolfgang Viereck (eds.), *Englisch, Formen und Funktionen einer Weltsprache*: *Ausstellung des Lehrstuhls für Englische Sprachwissenschaft und Mediävistik und der Universitätsbibliothek, Univ. Bamberg, 19. Mai - 30. Juni 1983*, 51–72. Bamberg: Lehrstuhl für Engl. Sprachwiss. u. Mediävistik.

Schneider, Edgar W. 2007. *Postcolonial English*: *Varieties around the world*. Cambridge: Cambridge University Press.

Schneider, Edgar W. 2011. *English around the world: An introduction*. Cambridge: Cambridge University Press.

Schneider, Edgar W. 2012. Exploring the interface between World Englishes and second language acquisition – and implications for English as a lingua franca. *Journal of English as a Lingua Franca* 1(1): 57–91.

Schneider, Edgar W. 2013. English as a contact language: The "New Englishes". In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language*, 131–148. Cambridge: Cambridge University Press.

Schneider, Gerold & Marianne Hundt. 2012. 'Off with their heads': Profiling TAM in ICE corpora. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide: Corpus-based studies of new Englishes*, 1–34. Amsterdam: Benjamins.

Schopf, Alfred. 1984. *Das Verzeitungssystem des Englischen und seine Textfunktion*. Tübingen: Niemeyer.

Schottman, Wendy. 1997. *Le Present Perfect face au colloquial Preterit en anglais américain et britannique contemporains*. Paris: Université de Paris 13 PhD.

Schreier, Daniel. 2010. The consequences of migration and colonialism II: Overseas varieties. In Peter Auer & Jürgen E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation. Theories and methods*, 451–467. Berlin: Mouton de Gruyter.

Schreier, Daniel. 2012. Varieties resistant to standardization. In Raymond Hickey (ed.), *Standards of English: Codified varieties around the world*, 354–368. Cambridge: Cambridge University Press.

Schreier, Daniel & Marianne Hundt, eds. 2013. *English as a contact language*. Cambridge: Cambridge University Press.

Schwenter, Scott A. 1994. 'Hot news' and the grammaticalization of perfects. *Linguistics* 3. 995–1028.

Scott, Mike. 2011. *WordSmith Tools*. Liverpool: Lexical Analysis Software.

Seoane, Elena & Cristina Suárez-Gómez. 2013. The expression of the perfect in South-East Asian Englishes. *English World-Wide* 34(1). 1–25.

Sempere-Martinez, Juan A. 2008. Cultural and psychological factors for the present perfect in British English and American English. *The International Journal of Arts Education* 6(1). 123–139.

Setter, Jane, Cathy S. P. Wong & Brian H.-S. Chan. 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.

Shannon, Thomas F. 1995. Toward a cognitive explanation of perfect auxiliary variation: Some modal and aspectual effects in the history of Germanic. *American Journal of Germanic Linguistics & Literatures* 7(2). 129–163.

Sharma, Devyani. 2001. The pluperfect in native and non-native English: A comparative corpus study. *Language Variation and Change* 13. 343–373.

Sharma, Devyani. 2009. Typological diversity in New Englishes. *English World-Wide* 30(2). 170–195.

Sharma, Devyani. 2012. Indian English. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The electronic world atlas of varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org/language/52> (17 September 2012).

Shields, Kathryn. 1989. Standard English in Jamaica: A case of competing models. *English World-Wide* 10(1). 41–53.

Sibayan, Bonifacio P. 1996. Post-imperial English in the Philippines. In Joshua A. Fishman (ed.), *Post-imperial English: Status change in former British and American colonies, 1940–1990*, 139–172. Berlin: Mouton de Gruyter.

Siemund, Peter. 2004. Substrate, superstrate and universals: Perfect constructions in Irish English. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 401–434. Berlin: Mouton de Gruyter.

Slobin, Dan I. 1994. Talking perfectly: Discourse origins of the present perfect. In William Pagliuca (ed.), *Perspectives on grammaticalization*, 119–133. Amsterdam: Benjamins.

Smith, Carlota. 1995. The range of aspectual situation types: Derived categories and a bounding paradox. In Pier M. Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality: Vol. 2. Typological perspectives*, 105–124. Torino: Rosenberg & Sellier.

Sokal, Robert R. & F. James Rohlf. 1962. The comparisons of dendrograms by objective methods. *Taxon* 11. 33–40.

Spolsky, Bernard. 2009. *Language management*. Cambridge: Cambridge University Press.

Strevens, Peter. 1992. English as an international language: Directions in the 1990s. In Braj B. Kachru (ed.), *The other tongue: English across cultures*, 27–47. Urbana: University of Illinois Press.

Stump, Gregory T. 1985. *The semantic variability of absolute constructions*. Dordrecht: Reidel.

Szmrecsanyi, Benedikt. 2012. Typological profile: L1 varieties. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 826–843. Berlin: Mouton de Gruyter.

Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. Vernacular universals and angloversals in a typological perspective. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular universals and language contacts: Evidence from varieties of English and beyond*, 33–53. London: Routledge.

Szmrecsanyi, Benedikt & Christoph Wolk. 2011. Holistic corpus-based dialectology. *Brazilian Journal of Applied Linguistics* 11(2). 561–592.

Tagliamonte, Sali. 1999. The Perfect and the preterite in contemporary and earlier English. By Johan Elsness. *Journal of English Linguistics* 27. 375–377.

Tagliamonte, Sali. 2000. The grammaticalization of the present perfect in English: Tracks of change and continuity in a linguistic enclave. In Olga Fischer, Anette Rosenbach & Dieter Stein (eds.), *Pathways of change: Grammaticalization in English*, 329–354. Amsterdam: Benjamins.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Thieroff, Rolf & Monika Budde. 1995. Are tense and aspect categories? In Pier M. Bertinetto, Valentina Bianchi, Östen Dahl & Mario Squartini (eds.), *Temporal reference, aspect and actionality: Vol. 2. Typological perspectives*, 47–62. Torino: Rosenberg & Sellier.

Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah G. 2010. Contact explanations in linguistics. In Raymond Hickey (ed.), *The handbook of language contact*, 31–47. Chichester: Wiley-Blackwell.

Thompson, Ellen. 2005. *Time in natural language: Syntactic interfaces with semantics and discourse*. Berlin: Mouton de Gruyter.

Thompson, Roger M. 2003. *Filipino English and Taglish: Language switching from multiple perspectives*. Amsterdam: Benjamins.

Tickoo, Makhan L. 1996. Fifty years of English in Singapore: All gains, (a) few losses? In Joshua A. Fishman (ed.), *Post-imperial English: Status change in former British and American colonies, 1940–1990*, 431–455. Berlin: Mouton de Gruyter.

Trudgill, Peter & Jean Hannah. 2008. *International English: A guide to the varieties of standard English*. London: Hodder Education.

Vaish, Viniti. 2007. Globalisation of language and culture in Singapore. *International Journal of Multilingualism* 4(3). 217–233.

Vaish, Viniti. 2008. *Biliteracy and globalization: English language education in India*. Clevedon: Multilingual Matters.

Van Herk, Gerard. 2008. Letter perfect: The present perfect in early African American correspondence. *English World-Wide* 29(1). 45–69.
Van Herk, Gerard, 2010. Aspect and the English present perfect: What can be coded? In James A. Walker (ed.), *Aspect in grammatical variation*, 49–64. Amsterdam: Benjamins.
van Rooy, Bertus. 2009. The shared core of the perfect across Englishes: A corpus-based analysis. In Thomas Hoffmann & Lucia Siebers (eds.), *World Englishes – problems, properties and prospects*, 309–330. Amsterdam: Benjamins.
van Rooy, Bertus & Lande Schäfer. 2002. The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies* 20. 325–335.
Vasudeva, Harender N. 1985. *Tense and aspect in English*. Ann Arbor: University Microfilms.
Vater, Heinz. 2007. *Einführung in die Zeit-Linguistik*. Trier: WVT.
Veloudis, Ioannis. 2003. Possession and conversation: The case of the category perfect. In Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.), *Perfect explorations*, 381–399. Berlin: Mouton de Gruyter.
Vendler, Zeno. 1957. Verbs and times. *The Philosophical Review* 66(2). 143–160.
Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.
Vermant, Stefan. 1983. *The English present perfect: A dynamic-synchronic approach*. Wilrijk: Universiteit Antwerpen.
Vlach, Frank. 1993. Temporal adverbials, tenses and the perfect. *Linguistics and Philosophy* 16. 231–283.
Wächtler, Kurt. 1977. *Geographie und Stratifikation der englischen Sprache*. Düsseldorf: Bagel.
Walker, Jim. 2008. The footballer's perfect – are footballers leading the way? In Eva Lavric, Gerhard Pisek, Andrew Skinner & Wolfgang Stadtler (eds.) *The linguistics of football*, 295–303. Tübingen: Narr.
Walker, Jim. 2011. The emergence of the narrative present perfect in British English: Reality or illusion? *Groninger Arbeiten zur germanistischen Linguistik* 53(2). 71–87.
Ward, John M. 1967. *The use of tenses in English*. London: Longman.

Wee, Lionel. 2004. Singapore English: Morphology and syntax. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie & Clive Upton (eds.), *A handbook of varieties of English*, 1058–1072. Berlin: Mouton de Gruyter.

Weinrich, Harald. 2001. *Tempus: Besprochene und erzählte Welt*. München: Beck.

Werner, Valentin. 2013a. The present perfect and definite temporal adverbials: Reference grammars and corpus evidence. *English Language Overseas Perspectives and Enquiries* 10(1). 9–21.

Werner, Valentin. 2013b. Temporal adverbials and the present perfect/past tense alternation. *English World-Wide* 34(2). 202–240.

Werner, Valentin. In preparation a. Overlap and divergence – aspects of the present perfect in World Englishes. In Elena Seoane & Cristina Suárez-Gómez (eds.), *World Englishes: New Theoretical and Methodological Considerations*. Amsterdam: Benjamins.

Werner, Valentin. In preparation b. The present perfect as a core feature of World Englishes. In Anna Mauranen, Markku Filppula & Juhani Klemola (eds.), *Selected papers from the Changing English 2013 conference*.

Wichmann, Søren & Matthias Urban. 2012. Toward an automatic classification of Englishes. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 676–686. Oxford: Oxford University Press.

Winford, Donald. 1993. Variability in the use of perfect have in Trinidadian English: A problem of categorical and semantic mismatch. *Language Variation and Change* 5. 141–187.

Winford, Donald. 2009. The interplay of 'universals' and contact-induced change in the emergence of new Englishes. In Markku Filppula, Juhani Klemola & Heli Paulasto (eds.), *Vernacular universals and language contacts: Evidence from varieties of English and beyond*, 206–230. London: Routledge.

Winford, Donald. 2013. Substrate influence and universals in the emergence of contact Englishes: Re-evaluating the evidence. In Daniel Schreier & Marianne Hundt (eds.), *English as a contact language*, 222–241. Cambridge: Cambridge University Press.

Wolf, Hans-Georg & Frank Polzenhagen. 2009. *World Englishes: A cognitive sociolinguistic approach*. Berlin: Mouton de Gruyter.
Wright, Susan. 1986. *Tense, aspect and text: Processes of grammaticalization in the history of the English auxiliary*. Cambridge: University of Cambridge PhD.
Wynne, Terence. 2000. *The present perfect: A corpus-based investigation*. Stirling: University of Stirling PhD.
Xiao, Richard. 2009. Multidimensional analysis and the study of World Englishes. *World Englishes* 28(4). 421–450.
Yao, Xinyue & Peter Collins. 2012. The present perfect in World Englishes. *World Englishes* 31(3). 386–403.

University
of Bamberg
Press

This study offers an account of the Present Perfect and its usage in worldwide varieties of English. It is based on corpus data from the International Corpus of English (ICE), which is assessed both quantitatively and qualitatively. It follows up on two broad desiderata in Present Perfect research, namely (i) to extend the knowledge of this grammatical area beyond the traditional British/American English paradigm, and (ii) to provide a systematic account of language-external effects such as modes of discourse, macro-genres and text categories. Occurrences of the Present Perfect are automatically extracted from part-of-speech tagged corpus files. Representative samples of the occurrences are then manually annotated for various factors (such as semantics, Aktionsart, temporal adverbials, sentence type, preceding tense, etc.) so that the distributions and the relative importance of these factors can be analyzed. In addition, the impact of alternative (non-standard) surface forms that may express a Present Perfect notion is considered. Measures of similarity between the various varieties of English under investigation are established (e.g. with the help of multidimensional aggregational methods such as cluster analyses and phylogenetic networks) and findings are related to a number of general models of World Englishes, whose descriptive adequacy for this particular area of grammar is tested.

eISBN: 978-3-86309-226-9

www.uni-bamberg.de/ubp