Non-canonical structures, they use them differently.

Information packaging in spoken varieties of English

Claudia Winkle

NIHIN studies
New Ideas in Human Interaction

Albert-Ludwigs-Universität Freiburg
Universitätsbibliothek

Non-canonical structures, they use them differently.

Information packaging in spoken varieties of English

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Philologischen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von

Claudia Winkle
aus Backnang

WS 2014/2015

Titel der eingereichten Dissertation:
World Englishes, do they use non-canonical sentence patterns?
Information-packaging constructions in spoken varieties of English

Erstgutachter: Prof. Dr. Dr. h.c. Christian Mair
Zweitgutachter: Prof. Dr. Dr. h.c. Bernd Kortmann
Drittgutachter: Prof. Dr. Peter Auer

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission der
Philologischen, Philosophischen und Wirtschafts-
und Verhaltenswissenschaftlichen Fakultät: Prof. Dr. Hans-Helmuth Gander

Datum der Disputation: 24.07.2015

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Introduction and research questions

Among the languages of the world, English can be seen as a language with a relatively fixed word order. Typically, sentences are organized around a subject and a verb in the canonical order of subject-verb-object (or complement or adverbial). Yet, sometimes speakers may deviate from this rather fixed word order for pragmatic purposes and shift elements in the clause into non-canonical positions in order to introduce new information into the discourse, re-introduce information at a later stage in the discourse, contrast one piece of information with another, or focus the addressee's attention on a certain piece of information. There are a number of syntactic devices which serve well for these purposes and which are particularly common in spoken interaction, where they reflect the dynamic use of language in a social context. The present study examines in some detail a number of such constructions, including left dislocation (e.g. *this cat she is fourteen*), right dislocation (e.g. *he is brilliant your dad*), fronting constructions (e.g. *my birthday party you arrange*), existential *there*-constructions (e.g. *there's a wild lot of people find her okay*) and various types of cleft constructions (e.g. IT-cleft: *it was you that told me that*; WH-cleft: *what he decided was to stay with some friends there*).

Previous studies have mainly focussed on the historical origins and the development or the discourse functions of the constructions in question (cf. e.g. Prince 1985; Aijmer 1989; Geluykens 1992; Biber et al. 1999; Gregory/Michaelis 2001; Netz/Kuzar 2007; Timmis 2010; Netz/Kuzar/Eviatar 2011; Reeve 2012; Patten 2012a). A number of more recent studies have added a cross-varietal perspective to the discussion by comparing the patterns of use across varieties of English. These studies claim, for example, that left dislocation and fronting constructions occur particularly frequently in the so-called 'New Englishes' [1](e.g. Platt et al. 1983: 14; Mesthrie 1992: 110; Bhatt 2004: 1023; Lange 2012: 148; Sharma 2012a: 214), and that IT-cleft constructions and fronting constructions show a high frequency of use and a wider scope of realization options in the so-called 'Celtic Englishes' [2] (cf. e.g. Kallen 1994; Filppula 1999; Hickey

1 The New Englishes will be discussed in some more detail in Chapter 2.

2 The term refers to the varieties of English spoken on the western edges of the British Isles and includes Irish English, Welsh English, Scottish English, Manx English (spoken in the Isle of Man) and Cornish English (spoken in Cornwall). The adequacy of this cover term is still being debated, yet

2007, 2012a; Beal 2012; Filppula/Klemola 2012). Furthermore, left dislocation has been identified as a common feature of learners of English in general (cf. e.g. Gruber 1967; Chambers 1973; Cotton 1978; Williams 1987; Carter/McCarthy 1995).

A notoriously difficult question is that of identifying possible explanatory factors for the observed usage patterns. This is particularly difficult in multicultural and multilingual societies because we often find a complex network of various interacting forces at play (cf. e.g. Thomason 2010; Sharma 2012a). One possible influencing factor is language contact, and especially so in the context of the New Englishes, where different languages permanently come into contact and where speakers have a repertoire of multiple languages at their disposal. In such cases, it is very likely that features of the substrate are being transferred into the contact language. But considering the typologically varying background languages in situations where English comes into contact with other languages, how can we, for example, account for similar trends among the New Englishes or for similarities between the Celtic Englishes and the New Englishes? So there must be other influencing factors as well. Possible candidates that have been suggested in the literature include general universals of human conceptualization and grammaticalization (e.g. processing and economy constraints, frequency, markedness; cf. e.g. Heine/Kuteva 2010; Diessel 2007), acquisitional universals (on the effects of second language acquisition on left dislocation cf. e.g. Williams 1987; Carter/McCarthy 1995; Ortega 2009) or the sociolinguistic and pragmatic setting. This last aspect is particularly relevant for multilingual settings, where English is spoken as only one language among (many) others, that is, where different languages, cultures and traditions come into contact.

All the studies that have been carried out on information-packaging constructions so far provide valuable insights into their distribution and use across different varieties of English. They also raise a number of questions, however. The features under consideration have been defined and labelled in different ways by different researchers and thus the (quantitative) findings and discussions include diverging structures, making them difficult to compare. Furthermore, quantifications such as 'very frequent' or 'quite frequent' are hard to assess properly without any numbers of comparison provided. The present study aims at contributing to this pool of research by systematically analyzing and comparing the use of left dislocation, right dislocation, fronting, existential *there*-constructions and cleft constructions across a number of first- and second-language varieties of English. With such a comprehensive survey it will be possible to directly compare speakers' preferences and properly assess the quantitative findings. Furthermore, it is expected that, by taking a cross-varietal approach, the present study will be able to identify qualitative properties which are unique to one or the other variety, that is, to uncover idiosyncratic features which are rather rare and thus difficult to observe. As for the possible explanatory factors for the observed usage patterns, the present study seeks to test previous claims and to offer new insights into the mechanisms of language contact and its outcomes or, more precisely speaking, of the interplay of language contact, universal learner and processing strategies and sociolinguistic and pragmatic factors.

Filppula (2006) claims that "the largely similar conditions of emergence of these dialects, combined with a number of shared linguistic features which have close Celtic parallels, lend enough support to the term Celtic English as a useful 'working concept'" (2006: 507).

The major research questions the present study thus addresses include the following: Do speakers of different English varieties show differences in the way they structure the information in a sentence, that is, do they have different preferences in the use of left and right dislocation, fronting, existential *there*-constructions and clefts? If they do so, are the differences quantitative or qualitative in nature? Are the structures used for the same purposes? If there are differences, which motivating factors can be identified?

1.2 Data and methodology

The study is based on two very different groups of corpora, one small and carefully-curated, the other "big and messy". The analysis of various components of the International Corpus of English (ICE) project is to provide a comprehensive picture of the constructions in question across varieties of English. One important limitation of the ICE corpora is their size, with the one-million-word corpora being small for today's standards. Because of this, some features analyzed are so rare in the ICE corpora that no in-depth analyses are possible and reliable conclusions can often not be drawn. Hence, the analysis of some features will be complemented by searches in two larger corpora, namely the Corpus of Contemporary American English, which contains 450 million words of speech and writing, and the Corpus of Global Web-Based English, a corpus of roughly 1.9 billion words based on web pages.

The one-million-word ICE corpora consist of spoken and written data.[3] Since the features under discussion are typical of spoken interaction rather than written language, the data used for the present analysis have been sampled from the 'private dialogues' sections, which comprise 90 texts of recorded and transcribed face-to-face interactions and 10 texts of transcribed telephone conversations. These 100 text files consist of about 2,000 words each, adding up to samples of about 200,000 words for each variety of English analyzed.

The present study rests on the assumption that the data provided by the ICE project are comparable across corpora and allow for comparative studies of English worldwide (as announced on the project's homepage). This is to be guaranteed by the common design of all ICE corpora, which means that they include texts from speakers and writers of similar profiles, use the same text categories and date from broadly the same period. Preceding the collection of the data for the ICE corpora, quite some effort was undertaken to discuss the common design of the corpora and to guarantee that only those text types were included that were applicable in all countries. Yet, it has to be noted that total compatibility cannot realistically be achieved, of course. For example, for some corpora the direct conversations were sampled in more formal settings than in others, which may skew the data in one way or another. The British English component in ICE, for example, includes conversations between students and their super-

3 Appendices 6.1 and 6.2 provide a concise overview of the genres included in the ICE corpora. For more information on the composition of the ICE corpora also see Greenbaum (1996), Kachru et al. (2009) and the ICE project's homepage at http://ice-corpora.net/ice/ (last accessed: Nov. 2015).

visors and between doctors and patients, while in ICE-New Zealand and ICE-Ireland it is mainly students and friends among themselves who have been recorded. These differing communicative situations obviously constitute different levels of formality. Furthermore, some of the direct conversation files in the British English corpus consist of interviews, also a more formal mode of discourse than interactions among friends. It can be expected that the grammatical features examined in the present study are not equally distributed across these spoken genres (even though all these genres are part of the corpora's private dialogue files). Hence, when analyzing and discussing the data these factors should be kept in mind since some of the observed usage patterns may well be due to this imbalance in data collection or genre classification.

Another issue that has to be kept in mind when analyzing the data concerns the situation of English in the different countries and the purposes it is used for. As has been mentioned above, the data have been sampled from the 'private dialogues' files which are meant to represent informal communication of the respective countries. Note, however, that in the countries where English is spoken as a second language most speakers would normally not use the English language for informal conversations but rather their national language or their home vernacular (e.g. Filipino/Tagalog or some minority language in the Philippines, Hindi or some other Indian language in India, Mandarin in Singapore, Jamaican Creole in Jamaica, Cantonese in Hong Kong). This is nicely expressed by a speaker from ICE-Philippines in the following way: "English is the common language to use in the class but uh if you're talking with fellow uh Filipinos then you can converse in Tagalog" (ICE-PHI:S1A-091). In some countries/territories, English may be used for specific purposes only, for example, in the government, law and international business, as in Hong Kong; in others, such as Singapore and India, it may additionally serve intra-ethnic communicative purposes or it may even be used at home (in Singapore, but rather the basilectal variant). Deuber (2009) discusses the problems of recording informal conversations for the Jamaican ICE component, a challenging but not infeasible task, as she notes:

> [...] although recording appropriate private interactions for Caribbean ICE corpora is certainly a challenge, fieldworkers have so far been able to strike a good balance between the demands of recording 'English' and recording 'conversations'. The recordings represent a range of language use, which is determined by a complex interplay of situational and social factors. (Deuber 2009: 432)

A further challenge involves the question of who counts as an "educated speaker of English". Of course, the candidates to be included in the corpora have been defined as having "received formal education through the medium of English to the completion of secondary school" (Greenbaum 1996a: 6). Still, people matching this definition in, say, the Philippines might have quite a different level of education than speakers from, for example, Canada or Singapore (See Platt et al. (1984: 164f.) on this problem.).

Further problems may arise in countries which have large proportions of immigrants. The collectors of the data have to confront the question of, as Janet Holmes puts it, "Who counts as a New Zealander?" (1996: 164f.). This problem appears to be particularly relevant for ICE-Hong Kong because Hong Kong's population consists of people from many different countries, as a speaker in the corpus also points out:

"there's nothing <,> special about Hong Kong <,> because is just a place <,> that uhm you know <,> consist of different country <,> I mean *different people from different country*" (ICE-HK:S1A-073) [emphasis mine].[4] Furthermore, the Hong Kong component contains many speakers who do not meet the conditions to be included in the corpus because they come from, for example, Japan. These speakers must, of course, be excluded from the analysis, but they still somehow impact on the conversation.[5]

The small sample size and the question of compatibility of the data are definitely important issues that seriously need to be considered. Yet, despite these drawbacks the ICE corpora are an invaluable tool for the present cross-varietal study since they provide language data of a large number of English varieties spoken around the world.

In total, spoken data from nine different varieties of English have been scrutinized. Since high frequencies of use of topicalization strategies have been attested for the New Englishes, the English varieties spoken in India, Hong Kong, Singapore, Jamaica and the Philippines have been included in the analysis. Furthermore, data from four first-language varieties (L1s) were added as points of reference. These include data from ICE-Great Britain, ICE-Ireland, ICE-Canada and ICE-New Zealand.[6] A comparison with British English is of particular interest because it is the input variety to all L2 English varieties but Philippine English, which derives from American English. Furthermore, it will be interesting to see whether Irish English, belonging to the group of Celtic Englishes and having developed in a similar way as the L2 English varieties, shows preferred usage patterns similar to those of the other L1 varieties or whether it behaves more like the L2 varieties analyzed.

Previous studies suggest that differing trends can be expected across varieties of English, the significance depending on the kind of construction and the variety. The present study will provide a fine-grained analysis of these tendencies. Following common practice in corpus-based investigations, it will not only report on quantitative findings but will also give qualitative, functional interpretations of the patterns of language use identified in the nine corpora. The limitations of a quantitative approach and the advantage of a qualitative approach are nicely summarized by Calude (2009a: 29-30) in the following words:

> [...] difficult judgements are often required to 'squeeze' elements into one category or another, when these clearly do not fit such 'smooth categorization' [...]. This results in an idealized model of the data and in some cases, in a somewhat empty and trivial one. [...] The advantage of a qualitative approach is that we are able to obtain a rich analysis and a detailed perspective of the data, while giving rare constructions just as much attention as frequent ones. This means that outliers or controversial cases are not missed, and that an idealized Aristotelian model [...] is not imposed.

I absolutely agree with Calude and the present study definitely profits from a combination of quantitative and qualitative methods, especially when the numbers get low

4 The symbol <,> indicates a short pause.

5 For more information about the problems of compilation also see, for example, Schmied (1996) and Mair (1992).

6 At the time of culling and analyzing the data, the Australian component of ICE was hosted at Macquarie University, Sydney, and accessing the text files was not possible (only searches via the web interface were possible). Since July 2014 the text files have been made available at https://www.ausnc.org.au/corpora/ice (after authorization and registration). Yet at that time, the present study was at such an advanced stage that it was decided to leave the analysis of ICE-Australia for future research.

and conclusive generalisations are difficult to make.

Another way of compensating for the limited size of the ICE corpora and verifying less robust findings is the use of larger corpora. Hence, the Corpus of Contemporary American English (COCA) and the Corpus of Global Web-Based English (GloWbE) have been used for some features as an additional tool. COCA contains 450 million words of American English dating from 1990 to 2012. The data are culled from five different genres, namely spoken, fiction, popular magazines, newspapers and academic journals. GloWbE is based on 1.9 billion words of text from 20 different English-speaking countries. The texts are taken from about 1.8 million web-pages consisting of informal blogs (about 60% of the corpus) and other, more formal web-material, such as newspapers, magazines and company websites.[7] GloWbE is used in the present study as an additional important tool for researching low-frequency items. It has to be noted, however, that the corpus is used only as an approximation to the ICE data because it does not contain spoken data. Yet, it is assumed that the language of informal blogs and discussion forums constitutes a written mode of discourse that is quite close to speech. GloWbE is thus used to find more conclusive evidence for or against usage patterns for which only very low frequencies can be attested in the components of ICE.

Data sampling in ICE has been done by thoroughly reading through the 100 sample texts of all nine corpora and manually annotating the constructions under consideration, based on the definitions given by two standard grammar books (Quirk et al. 1985; Biber et al. 1999) and other relevant literature on the topic. Obviously, this was a rather time-consuming and at times daunting task, but it was definitely worth the effort. I could not and still cannot think of any other strategy (e.g. some kind of automated search) which would allow me to cull left and right dislocation and fronting constructions from the texts in a similarly exhaustive way. Furthermore, reading through all the texts gave me the opportunity to familiarize myself more closely with the speakers in the relevant corpora and the type of conversations they conducted.

As required by the common ICE design, the sample size of each corpus amounts to about 200,000 words. Since some variation can be expected and in order to improve the compatibility of the results of the analysis, the word counts of all samples have been computed with the help of the open source software R.[8] The resulting sample sizes range from 201,645 to 237,974 words.

COCA and GloWbE have been accessed via the search interface provided on the web-pages of the corpora.

7 GloWbE and COCA are freely available to all researchers at http://corpus.byu.edu/ (last accessed: Nov. 2015). A detailed overview of their make-up is given in Appendices 6.3 and 6.4.

8 R is available at http://www.r-project.org/ (last accessed: Nov. 2015). The word counts of the different ICE samples and the R code with which they have been computed can be found in Appendix 6.5.

1.3 Determinants of emergent language structure

In the present study, grammar is seen as a dynamic system that is constantly changing under the influence of multiple internal and external forces. As noted earlier, these factors usually interact to a greater or lesser extent in the emergence of linguistic knowledge in the different language settings. These interacting forces and some basic assumptions underlying the present study will be discussed in some more detail in the following paragraphs.

First, particular communicative and cognitive pressures of language use can impact on the emergence of linguistic structure, as suggested by usage-based linguists. Usage-based approaches share the basic assumption that linguistic structure is shaped by language use and that speakers' linguistic knowledge is based on past experiences (cf. e.g. Bybee 2007; Bybee/Hopper 2001; Diessel 2007, 2011; Bao 2010). For the present study it is important to note that past experiences also include other languages than English, that is, the languages learned alongside or before the acquisition of English. In the usage-based model, frequency of occurrence plays an important role as it affects the processes of language acquisition, sentence comprehension and processing, and diachronic change. For example, if a word or construction is frequently used, its representation in the speaker's memory is strengthened, that is, it is more entrenched; or if certain elements are arranged in recurrent orders, this may raise expectations in the speaker/hearer as to which element may occur after a certain expression; or if certain expressions are frequently combined, they may become automatized chunks, that is, linguistic units which are stored as a whole in the speaker's memory (cf. Diessel 2007). These effects of frequency of occurrence have a number of consequences for the present study, resulting in the following hypotheses. First, if speakers of one English variety use a marked structure[9] more frequently than speakers of other varieties, this structure is more entrenched in these speakers' memories and may eventually become an unmarked structure or at least less marked. The present study shows, for example, that left dislocation and fronting constructions seem to be less marked for Indian English speakers than for the speakers of the other varieties. Additionally, the entrenchment of a construction in the speakers' memories may result in an even more frequent use of the construction and a widening of the scope of realization options. Examples of this phenomenon can be found in the speech of Irish English speakers. The speakers of this variety tend to use left dislocation and IT-clefts not only more frequently than the other L1 speakers but they also show more variation, as the following analyses will show.

In addition to these features relating to frequency of use there are other principles that can affect the emergence of linguistic structure. Other cognitive principles, for example, include analogy and related phenomena (e.g. metaphor), which have been identified as important factors in both language acquisition and language change (Diessel 2007: 124). This also includes situations of L2 acquisition, that is, analogy across different languages or L1 transfer (cf. e.g. Ortega 2009), also referred to as

9 The features under discussion may be termed 'marked structures' since they transform in one way or another the canonical or unmarked word order of English clauses.

'interlingual identification' (Weinreich 1968: 7). With respect to syntactic structures this means that bilingual or multilingual speakers establish replica patterns in one language on the model of another language of their linguistic repertoire (cf. e.g. Matras 2009). This second factor affecting the emergence of linguistic knowledge - transfer of features from another language - applies to both language contact situations, where speakers grow up acquiring two or even more languages simultaneously, and English L2 settings, where English is learned in late childhood, adolescence or adulthood after the acquisition of another first language or other first languages.

It is well-known from the extensive body of literature on language contact that various interacting factors, both linguistic and extra-linguistic ones, are involved in the shaping of the grammar of a contact language (cf. e.g. Weinreich 1968; Thomason/Kaufman 1988; Thomason 2001; Siemund/Kintana 2008; Ansaldo 2009; Matras 2009; Hickey 2010, among many others). One important assumption underlying the present study is that the properties of a new variety can be understood in terms of the selection of a number of structural features from a feature pool to which the languages in contact contribute (Mufwene 2001, 2008; Ansaldo 2009; Gisborne 2009). That is, the features of different languages in contact are in competition and speakers have a certain degree of choice as to which of the competing (phonological, morphological, lexical or syntactic) features to use in a certain context. These choices may be conscious, for example, in situations where one feature has more overt prestige than the others, but they may also be unconscious and "depend on matters of cognitive salience, typological dominance as well as frequency" (Ansaldo 2009: 135).

Research in second language acquisition (SLA) and contact linguistics suggests that it is not only speakers' selections from the structural features available to them in their feature pool that are responsible for the shaping of a new variety but that universal developmental forces can also have an impact. Such forces may lead to similar results in different multilingual settings with typologically very different background languages. For example, for left dislocation constructions it has been claimed that they are found particularly frequently in the speech of learners of English - no matter what background languages they speak and including children who learn English as their L1 (e.g. Williams 1987; Carter/McCarthy 1995). Other researchers see the constructions as a L1-influenced phenomenon, at least in settings with a topic-prominent background language (Ortega 2009: 45). The results of the present study suggest that left dislocation tends to be an acquisitional phenomenon indeed, but since there is also some variation among the learner varieties, it seems that some ecologies are more favourable than others (due to influence from the background languages).

Finally, pragmatic factors and the social and cultural setting can also affect the shaping of linguistic structure. As mentioned earlier, this is particularly relevant in multilingual societies where different languages, cultures and traditions come into contact. Important aspects that should be taken into account are, among many other things, the attitudes towards the languages involved, the manner of learning each language, the relative proficiency in each language and the typology of the languages in question.

This list of factors impacting on the emergence of linguistic knowledge is certainly not exhaustive and the features are, of course, not to be seen as isolated constraints. It will be interesting to examine in how far these factors interact, especially in situations

of language contact. This will be done in quantitative and qualitative terms, as noted earlier. Frequency of use is a measure that can easily be computed and compared across varieties of English, especially since the techniques for recording, storing and analyzing spoken language data have developed and improved. It is certainly true that frequency of use has an impact on cognitive representations, that is, frequency is the CAUSE of certain developments in linguistic structure, as has been illustrated above. In addition, it is interesting to explore in how far frequency of use can be seen as an EFFECT. This involves going beyond the mere reporting of quantitative findings and seeking to find explanations for why certain constructions occur more frequently in certain varieties of English. Likely candidates are the factors mentioned above.

1.4 The scope of the present study

Approaching information structure from a cross-varietal perspective, the present study aims at finding evidence of systematicity and rule-governed processes in the emergence and use of left dislocation, right dislocation, fronting constructions, existential *there*-constructions and cleft constructions in various L1 and L2 English varieties. Diverging patterns of use can be expected, both in quantitative and qualitative terms. Based on the assumptions outlined in section 1.3, the present study will examine the following hypotheses. The expected differences in frequency of use can be seen as both the cause for the emergence of linguistic structure and as the effect of other influencing factors on language use. Frequency of use may lead to the strengthening of certain features in the speaker's memory and a widening of scope of realization options. Factors impacting on the frequency of use involve substrate influence in language contact situations, universal developmental processes in language acquisition, and specific features of the socio-cultural setting. It is expected that speakers have different motivations for their preferred patterns of use, depending on the construction in question and the language setting. Furthermore, it is expected that a single cause can rarely be identified and that instead it is often a multiplicity of influencing factors that interact in the shaping of linguistic knowledge. If we find common trends in the L2 English varieties in comparison to the L1 varieties, this might be due to universals of L2 acquisition. If we find variation across various L2 varieties, there might be different explanatory options. First, the varieties in question might be at different developmental stages. For example, Singapore English is developing towards an L1 variety for many speakers in the territory, while in Hong Kong the English language lost some of its significance after the transfer from British to Chinese sovereignty in 1997.[10] Second, L1 transfer might lead to diverging usage patterns and speaker's preferences and, depending on the proficiency level, the influence may be more or less significant. Since the present study examines information structure,

10 For more information on the history of English and its current status in these countries see section 2.2.

the typological differentiation between subject-prominent and topic-prominent languages is of special interest. If this typological difference has indeed an impact on the constructions under discussion, we might expect to find similar trends in Singapore English and Hong Kong English since both varieties have Chinese dialects as their major background languages. It should be kept in mind, however, that L1 transfer is neither everywhere nor inevitable - the features of the L1 are not all equally susceptible to transfer - and that different acquisition outcomes for the constructions under consideration can be expected (even within one English variety). Third, variation across L2 varieties might be due to the socio-cultural setting. For example, the government in Singapore believes in the importance of the English language for the economic and social advancement of the nation and thus fosters an English-medium education system. This in turn leads to a higher proficiency level. The analysis and discussion of the data will show in how far these different scenarios can be attested for information-packaging constructions. It can be expected to find a complex system of interacting forces, but it is hoped that at least the strongest and most significant trends can be identified.

The structure of the present study is as follows. Chapter 2 first discusses a number of models of World Englishes. This will be followed by a description of the historical development and current status of English in Ireland, New Zealand, Canada, Singapore, the Philippines, Jamaica, India and Hong Kong, including the arrival of the first English-speaking people, the manner of learning the English language, governmental intervention in language planning, general attitudes towards English (prestige), among other things. Chapter 3 begins with an outline of the main concepts of information structure in general and a review of some relevant literature. This will be followed by an introduction of the constructions under consideration in the present study and their realization variants, including a discussion of previous research on the constructions. The chapter will close with a brief outline of information packaging strategies in some of the background languages, including a brief discussion of possible counterparts to the English constructions examined in this study. Chapter 4 will provide a detailed quantitative and qualitative analysis of the constructions under consideration, followed by a concluding discussion in Chapter 5. Additional material will be provided in the appendix in Chapter 6.

CHAPTER 2

World Englishes

This chapter is about the global spread of English and the classification of World Englishes. The first section discusses theoretical models which aim at capturing the diversity of the English language around the world. Furthermore, the major aspects of Schneider's (2003, 2007) Dynamic Model of the evolution of postcolonial Englishes will be summarized. The second section will deal with the varieties of English that are subject to the present study. The historical origins and the current situation of the English language will be outlined.

2.1 Models of World Englishes

The legacy of colonial Englishes has resulted in the existence of several transplanted varieties of English, the 'new varieties' of English or 'New Englishes'. They have emerged in former colonial territories (e.g. India, Singapore or the Philippines) where the English language has been retained after independence as an official language and has come into contact with indigenous languages. English in these countries is mainly used in administration, education, literature and the media. It is spoken as a mother tongue by only a small proportion of the population but is typically learned as a second (L2) or third (L3) language in educational institutions (cf. e.g. Platt et al. 1984; Foley 1988; Mukherjee 2007; Bao 2010).

The global spread of English has led to a growing interest in the identification and description of the new varieties of English in the Caribbean, West and East Africa and many parts of Asia. Especially during the 1980s and 1990s, many linguists proposed various models aiming at capturing the diversity of the New Englishes and classifying them as varieties of English in their own right. Additionally, models have been proposed that describe the developmental stages of a newly emerging variety. Among the most influential models are Kachru's (1982, 1988) three-circles model and McArthur's (1987) wheel model.[1] Kachru's model depicts the varieties of English spo-

1 Further models identifying and describing World English(es) have been suggested, for example, by Görlach (1991) and Melchers and Shaw (2003). A very comprehensive discussion of different approaches to World Englishes since the 1960s is given in Bolton (2003). The monograph also includes

ken around the world in terms of three concentric circles which "represent the types of spread, the patterns of acquisition, the range of functional domains, and the societal penetration of the language" (1988: 5). He names these three circles the 'Inner Circle', the 'Outer Circle' and the 'Expanding Circle'. The 'Inner Circle' refers to native-English-speaking countries such as the UK, USA, Australia, New Zealand and Canada, that use English as their primary language. These are norm-providing or endocentric. The 'Outer Circle' includes former British and American colonies such as India, Singapore and the Philippines, countries in which English has official status and functions (e.g. in administration and law) and is spoken as a second language. These countries are called norm-developing because English has already undergone some acculturation and nativization and there is a creative literature written in the local variety of English. The third circle, the 'Expanding Circle', refers to countries where English has the status of a foreign language, such as China, Indonesia, Japan and Korea, and its use is restricted to limited domains (e.g. education, international communication).

Kachru's model can be criticized for not allowing precise classifications. Some countries, for example, have Inner Circle and Outer Circle populations, such as Singapore and South Africa. Yet other areas scratch the line between Outer Circle and Expanding Circle. Hong Kong is a case in point. English in Hong Kong has the status of an official language and is widely used in the education system, just like in institutionalized Outer Circle varieties. However, like in countries of the Expanding Circle, its standards tend to be exonormative, and it is used for international purposes rather than in more informal situations. Jamaica constitutes another problematic case as it does not fit neatly into any of the three circles because of the complexity of its sociolinguistic situation.[2]

McArthur's wheel model does also consist of three parts, with World Standard English constituting the centre of the wheel. Outside the centre there is a circle of eight national and regional varieties, differentiated into established standard varieties and varieties which are still in the process of standardizing: British and Irish Standard English; American Standard English; Canadian Standard English; Caribbean Standard English; West, East and South(ern) African Standard(izing) English; South Asian Standard(izing) English; East Asian Standardizing English. Around this inner circle we find another circle which includes the national varieties within these eight regions (e.g. Welsh English, Singapore English, Hong Kong English), subnational varieties (e.g. Inuit English, Quebec English) and other subvarieties (e.g. BBC English).

What both these models have in common is the idea of nativization in the outer-circle varieties. This means that speakers of these English varieties have adopted a once foreign language and have adapted it to their cultural contexts. The notion of nativization has received particular attention because it "bridges the gap between the norm-producing inner circle and the norm-developing outer circle and because it has helped to establish New Englishes as full-fledged varieties besides the native varieties

extensive information on further literature on the topic. Mesthrie and Bhatt (2008: 27-36) discuss different models, their strengths and shortcomings, and suggest further models. Crystal (2003) gives a concise description of the global spread of English, including maps and numbers of English speakers in different countries.

2 Also see Mesthrie (2008) on the blurring of the circles in the twenty-first century.

of Englishes" (Mukherjee 2007: 160).

More recently, Schneider (2003, 2007, 2014) has suggested a dynamic model of the evolution of new varieties of English. Schneider's model rests on the assumption that

> a fundamentally uniform developmental process, shaped by consistent sociolinguistic and language-contact conditions, has operated in the individual instances of relocating and re-rooting the English language in another territory, and therefore it is possible to present the individual histories of PCEs [Postcolonial Englishes] as an instantiation of the same underlying process. More specifically, it is posited that evolving new varieties of English go through a cyclic series of characteristic phases, determined by extralinguistic conditions. (2007: 5)

Two factors, in particular, constitute the core of this developmental process. These factors are the reconstruction of identities and the changing interrelations between the settlers and the indigenous population. That is, due to prolonged contact on shared territory settlers and indigenous people gradually approximate in terms of culture and language, a process during which they rewrite their identities and assume new hybrid identities and new linguistic norms (Schneider 2007: 6). In line with Mufwene (2001, 2008), Schneider's model is based on the idea of a 'feature pool' of linguistic features from which settlers and indigenous people select and thereby reconstruct their linguistic and social identities (2007: 21).

The Dynamic Model assumes that emerging varieties proceed through five consecutive phases: foundation, exonormative stabilization, nativization, endonormative stabilization and differentiation.

In the foundation phase, English is established in a previously non-English-speaking territory. The settlers may be from different regions, which then may lead to dialect contact and koinéization. Among the indigenous people only those who are in contact with the settlers acquire some English.

With the stabilization of the territory as a colony in the phase of exonormative stabilization, the contact between the settlers and the indigenous people increases. The settlers' language is norm-providing, but lexical loans from the indigenous languages and some early phonological and syntactic transfer phenomena are found.

The phase of nativization is most central to the Dynamic Model because it is in this phase that the settlers and the local people begin to construct a new identity. Furthermore, the contact between the settlers and the local people increases further and the social gap between them gets smaller, that is, the interactions between the two groups get more intertwined. Through the increased amount of interaction a new variety of English emerges with its very own distinctive features in phonology, lexis and syntax. These features may be traced back to second language acquisition, L1 transfer and/or creative innovation.

The phase of endonormative stabilization is usually reached after political independence. Both settlers and local people increasingly identify with the newly formed nation and accept indigenous norms, that is, they no longer orient towards the settlers' original norms (typically British norms). The transition from phase three to phase four may additionally be marked or speeded up by an 'Event X' - "some exceptional, quasi-catastrophic political event" (Schneider 2003: 250). Furthermore, phase four shows attempts of codifying the local norms in dictionaries and grammar books and an increasing use of English for creative writing. It is also in this phase that the

expression 'English in X' is replaced by 'X English', a label which underlines the acceptance of the local norms and the identification with it.

The phase of differentiation is no longer concerned with the formation of a new nation but with the formation of subnational group identities and the emergence of new dialects and sociolects within the new variety.

Note that the Dynamic Model is, of course, only an abstract and idealized account of an evolutionary pattern that may underlie the formation of New Englishes. It does not depict reality itself. Furthermore, note that the phases have no clear-cut boundaries but rather shade into each other with features of different phases possibly even coexisting at some point during the developmental process.

The model has been applied to several case studies of Inner and Outer Circle countries, as will be seen in the following subsections, which deal with the origins and current situation of English in Ireland, Canada, New Zealand, Singapore, the Philippines, Jamaica, India and Hong Kong. These countries have their very own linguistic ecologies with many different factors having contributed and still contributing to the shaping of the relevant varieties of English because, as Gonzalez (2008) rightly points out,

> [w]hat becomes of the transplant is very much a function not only of geography but of the society which receives the transplant, which includes the types of languages already in use in the receiving country, the role if any of *linguae francae*, the role if any of the national language [...], the means by which the new transplant is propagated, and the social and economic dominance of the language. (Gonzalez 2008: 24)

A number of these factors will be discussed in the following subsections. Additionally, section 3.4 will introduce the major background languages (i.e. Irish, Mandarin, Cantonese, Tagalog, Jamaican Creole, Hindi and Malayalam) and their means of structurally marking emphasis in the sentence. This is meant to help identify and explain transfer phenomena and to better understand the innovative structures speakers may use.

2.2 The English language in different countries

2.2.1 Ireland

> *Linguistically, present-day IrE can be characterised as one of the 'Inner Circle' or 'L1' Englishes [...] From a socio-cultural and historical point of view, IrE can be described as one of the L2 varieties, as it has evolved as a result of long-standing coexistence and contacts with the indigenous Celtic language of the Irish people, Irish.*
> (Filppula 2012: 30-31)

The English language was first introduced into Ireland in the twelfth century, when

Anglo-Normans, English, Welsh and Flemish invaded the country (Hickey 2007: 30).[3] At first, it had to compete not only with Irish, the vernacular tongue of the population, but also with Latin and French, which were the languages of administration, education and law in both Ireland and England at that time. Since the majority of the population continued to use Irish, the English language began to go into decline as a spoken vernacular and by the end of the sixteenth century English-speaking people in Ireland were almost entirely assimilated to the Irish language and culture (Filppula 2012: 31). Contemporary reports nicely illustrate how the Anglo-Irish at that time perceived the loss of English as a vernacular and of the English way of life: "many English [...] forsaking the English language, fashion, mode of riding, laws and usages, live and govern themselves according to the manners, fashion, and language of the Irish enemies" (Statutes of Kilkenny, 1366; quoted in Kallen 1994: 152).

The English language survived in some of the major cities like Dublin and in few scattered rural areas in the east and south-east of Ireland. In these places, features of early dialects of English have thus been preserved up until the nineteenth century (cf. Kallen (1994: 167) and Hickey (2007: 66ff.) on the Forth and Bargy dialect).

In 1541, Ireland was incorporated into the Kingdom of England with the proclamation of Henry VIII as King of Ireland. This did not automatically mean that Irish was replaced by English overnight, but later in the century the tide indeed was to rise in favour of English (King 2006: 37). Queen Mary and, even more vigorously, James I instituted plantations, the settlement of English-speaking people in Ireland, which had the effect that further varieties of English were introduced, notably Scots in Ulster and various other English dialects generally. Although Irish was thereby pushed into more isolated areas, particularly into the south and west, it hold up its position remarkably well up to the end of the eighteenth century (Kallen 1994: 156; King 2006: 37; Filppula 2012: 31). Yet during the nineteenth century, the number of Irish speakers rapidly declined, with many people shifting to English and thus abandoning their native language. Different factors contributed to this process of language shift, which "proceeded at a pace scarcely paralleled in linguistic history" (Filppula 2012: 31).

One of the major factors was the Great Famine (1845-1848), which resulted in an decrease of the Irish population by two million due to death and emigration (McCartney 1987). Those who sought work in North America or Europe were for the most part rural inhabitants from the west and south of the country, that is, most of them were native speakers of Irish (Hickey 2007: 47). But even before the Famine the English language had already made great inroads into the Irish-speaking community. Daniel O'Connell, the leader of the Catholic Emancipation movement, was decidedly against the Irish language and chose English as the language of his campaign for utilitarian purposes, saying in 1833:

> [...] although the Irish language is connected with many recollections that twine around the hearts of Irishmen, yet the superior utility of the English tongue, as the medium of modern communication, is so great, that I can witness without a sigh the gradual disuse of the Irish. (quoted in Crowley 2000: 153)

But even within the wider Irish-speaking community Irish was willingly abandoned. With the establishment of the national school system in 1831 English was intro-

3 Only a brief sketch of the history of Irish English is provided here. For a more extensive discussion see, for example, Kallen (1994) and Hickey (2007).

duced as the medium of instruction. The Irish people, priests and political leaders co-operated willingly with this system because, although being the language of the colonizers, English offered benefits and promised social advancement if adopted (McCartney 1987). Parents even encouraged their children to learn English because it was seen as "the key to the golden door of America" (de Fréine 1977: 86).

The way the Irish acquired the English language is often described as 'unguided adult language acquisition' or 'group second language acquisition' because there was little if any formal education for the majority of the population. This uncontrolled and non-prescriptive way of acquisition provided the grounds for influence of the Irish language on Irish English[4], or as Winford (2003) puts it:

> The persisting of bilingualism within the shifting group is another important factor in language shift [...] there were large numbers of illiterate bilinguals in nineteenth-century Ireland, judging from the figures of the 1851 census. It is reasonable to assume also that childhood bilingualism was quite common, and that bilingual children played a role in the regularization of Irish English grammar. These factors would have favoured the retention of Irish features in the English of such speakers. (Winford 2003: 253)

Today, Irish survives only in three regions on the western seaboard, also known as the 'Gaeltacht': in the south-west, the mid-west and the north-west (Hickey 2007: 48; Filppula 2012: 31). However, it is recognized as the first official language in Article 8 in the Constitution of Ireland, while English is "recognised as a second official language" (Constitution of Ireland 2013: 8). Furthermore, governmental efforts to restore the Irish language have led to the situation that it is now widely studied and used as a second language by almost everybody who has gone through the educational system.

As for the attitudes towards Irish English, it should be pointed out that it is denied recognition as a variety of English in its own right even among many Irish even today. Irish English is often regarded as simply a substandard language not to be taken seriously (Hickey 2007: 23). Addressing the question of why the Irish do not hold their specific variety of English in higher regard, Hickey suggests that various factors be taken into account. Among these are linguistic prejudice against prominent features of Irish English and a feeling of inferiority of "anything homegrown", a post-colonial attitude "which still lingers on" (Hickey 2007: 23). Furthermore, acknowledging the language of the former colonizers is regarded as a kind of disloyalty to Irish. Hence, although English is the native language of the vast majority of the Irish population, many Irish have an ambivalent attitude to English.

A final issue that is worth mentioning is that today the shaping of Irish English is also influenced by large numbers of immigrants, coming in particular from Poland and Lithuania. The Census figures of 2011 show that the proportion of the non-Irish population now accounts for about 12% in the Republic of Ireland and 2.9% in Northern Ireland (Central Statistics Office 2014: 39).

4 Throughout the present study, 'Irish English' is used as a cover term to refer to both the variety of English spoken in the Republic of Ireland and the variety spoken in Northern Ireland. If a distinction is necessary it will explicitly be pointed out. Other terms found in the literature include 'Anglo-Irish' and 'Hiberno English' for the English variety in the Republic of Ireland. The forms of English in Northern Ireland are also referred to as 'Ulster English(es)', 'northern Irish English' or 'northern Hiberno English', among others. See Hickey (2007: 3ff.) for a detailed discussion of terminology.

2.2.2 New Zealand

> A: *In the North Island every sentence seems to finish with* eh [...]
> B: [laughs] *No that's not true* [...] *I don't talk like that* [...] *No there are certain people who do say that you're quite right yes*
> A: *And you know it's er I think it comes a lot from the Maori the system of er speaking*
>
> (ICE-NZ:S1A-100)

When the Europeans discovered New Zealand in the seventeenth century the country was already settled by the Maori, an Eastern Polynesian people, who probably travelled by canoe from somewhere in the tropical Pacific. The time of the arrival of these Polynesian people in New Zealand is uncertain. While some researchers say that they reached the country more than 1,000 years ago (Bauer 1994: 382; Hay et al. 2008: 3), others give the thirteenth century as the time of their arrival (Wilson 2014: 1).

The first Europeans to set foot on New Zealand were the Dutch Abel Tasman and his crew. On an expedition for the Dutch East India Company, he sighted 'a large land, uplifted high' - probably the Southern Alps - in 1642 (Wilson 2014: 2). Tasman called the land he had discovered 'Staten Landt', thinking that it was part of Australia. The name New Zealand, or rather 'Nieuw Zeeland', was coined by a Dutch cartographer to the Dutch East India Company later in the seventeenth century; the "only linguistic result" of the Dutchmen's discovery (Bauer 1994: 382).

James Cook's *Endeavour* was the first English ship that reached New Zealand, landing at Poverty Bay more than one hundred years after Tasman in 1769. He circumnavigated the island, thoroughly mapping the outline of its coast and providing Europe with the first comprehensive visual and written record of the country's nature and substantial knowledge of the Maori people (Wilson 2014: 6-7). Yet linguistically, Cook left no direct traces because "[w]hen the *Endeavour* left New Zealand, English left with it and did not take root in the country until the first European settlements at the end of the eighteenth century" (Kuiper/Bell 2000: 11).

From about 1792 onwards, sealers and whalers visited and operated from the coasts of New Zealand, but hardly any of them settled on the island. At that time, New Zealand was rather unappealing to Europeans: the journey was long and expensive, and the country was associated with the convict settlements of Australia and had a reputation as "a home of bloodthirsty cannibals" (Phillips 2013: 4-5).

In the late 1820s the number of non-Maori living in New Zealand began to rise, but by 1838 there were still only about 2,000 immigrants. This number was to increase decidedly in the following years resulting in a non-Maori population of about 10,000 people by 1842 (Bauer 1994: 383). There are two decisive factors for this change. First, in 1840 the British Government and Maori chiefs signed the Treaty of Waitangi, ceding sovereignty to the Queen and creating "the foundation for British colonial rule in New Zealand, which in turn created the framework for sustained migration from Britain" (Kuiper/Bell 2000: 12). Second, in the same year the first European settlers, assisted by the New Zealand Company, arrived in the country, bringing in many more immigrants in the following years (Bauer 1994: 383; Phillips 2013: 4).

Immigration added greatly to New Zealand's population in the following decades. As the census figures presented in Table 2.1 show, the population of 1886 is more than five times the size of that of 1861 (Statistics New Zealand 2013). Immigrants from the British Isles and in particular England made up the highest proportions. Yet by 1886, there were more New Zealand-born Europeans in New Zealand than immigrants (51.9%; cf. Table 2.1). Bauer (1994: 386) suggests that this date be taken "as a point after which the development of the English language in New Zealand reflected New Zealand rather than British or Australian trends".

The census figures show that most of the early immigrants to New Zealand came from the British Isles, but there is hardly any linguistic information about those early settlers. In many cases the only information available is about their port of embarkation, but this tells us little about the places where they were born or where they lived. This means that the origins of New Zealand English are difficult to exactly pin down and several explanations have been suggested in the literature (Bauer 2000).

When New Zealand English was first recognized as a new variety of English (around 1900), many people all around New Zealand began to complain that children were speaking with a 'colonial twang'. The most common explanation at that time was that New Zealand English was a transported variety of Cockney, a London dialect. But there is demographic evidence against this position, with only a small proportion of the early settlers actually coming from London (15%; cf. Hay et al. 2008: 85). Furthermore, Bauer points out that what is known of the Londoners' social class suggests "not only that they were not Cockneys, but that they would have despised a Cockney accent" (1994: 421). Given this evidence, Hay et al. (2008: 85) propose that at that time 'Cockney' may have been used as a general term of abuse rather than as an accurate description of the linguistic features of this variety of English.

Table 2.1: Proportion of persons of different birthplaces living in New Zealand (exclusive of Maoris) at the various census periods.

census year	**1861**	**1867**	**1874**	**1881**	**1886**
New Zealand	27.9	29.3	41.0	45.6	51.9
Australia	2.6	5.2	4.5	3.5	3.0
England	36.5	30.0	24.8	24.3	21.7
Wales	0.5	0.6	0.5	0.4	0.3
Scotland	15.7	15.9	12.9	10.8	9.5
Ireland	8.9	12.8	10.1	10.1	8.9
other British dominions/at sea	2.3	2.1	1.3	1.1	0.9
foreign countries	2.7	3.8	4.7	4.0	3.4
unspecified	2.9	0.4	0.2	0.2	0.4
total population	**99,021**	**218,668**	**299,014**	**489,933**	**578,482**

SOURCE: Statistics New Zealand, 1886 Census results, Table IV_II.

A second theory holds that New Zealand English could have been an exported version of Australian English. Proponents of this view give as evidence the overwhelming phonetic and phonological similarity between the two varieties of English and the

large number of words they have in common, "virtually to the exclusion of the rest of the English-speaking world" (Bauer 1994: 425-427). Furthermore, from early on there have been and still are close (economic) connections between both countries. New Zealand even started out as a colony of New South Wales, gaining independent colonial status in 1841 (Hundt 2012: 1997). What weakens the plausibility of this position, however, is the fact that the proportions of Australian immigrants to New Zealand have always been relatively small, as the census figures above show. Hence, Hay et al. (2008: 86) conclude that an Australian 'language planting' explanation can be ruled out, but they add that there was definitely an Australian influence, as many of the early settlers to New Zealand came via Australia.

A third explanation for the origins of New Zealand English states that it has not been transplanted from some other place but has developed independently in New Zealand itself. This process is also referred to as 'new-dialect formation' in sociolinguistics and explained by Hay et al. (2008) as follows:

> The theory is that when people come to a new country or a new region speaking different dialects, over time the different dialectal variants become levelled out and a single new dialect develops, which is different from those dialects that the first settlers used. (Hay et al. 2008: 86)

As has been noted above, those complaining about the 'colonial twang' commonly associated it with the speech of children. Studies of language change have shown that children and adolescents play an important role as agents of change in the process of new-dialect formation (cf. e.g. Kerswill/Williams' study (2000) in Milton Keynes). And this might have also been the case in New Zealand. The early population of New Zealand was a very young one, with one quarter of the population in the 1840s and 1850s being children, while there were few people over 45 (Hay et al. 2008: 93). With the Education Act of 1877 primary education in New Zealand became compulsory, leading to increasing numbers of children coming together for their education. Given this situation, it is not at all surprising that "the development of the New Zealand accent seems to have occurred and spread very rapidly in the 1880s" (Hay et al. 2008: 93).

A fourth theory argues in terms of a combination of the theories just mentioned. Examining various explanations for the origins and development of New Zealand English, Gordon et al. (2004) conclude that multiple factors contributed to the shaping of the language. These include input from immigrants who came via Australia and swamping effects from large-scale immigration in the 1870s. They add that they cannot say for sure "to what extent factors such as education, standardisation, and acts of identity may have influenced the final outcome, but neither can be eliminated; that we cannot confirm them does not mean that they had no effect" (2004: 258).

While immigration is still a great issue in New Zealand today - most immigrants are now coming from Asia and the Pacific Islands - the European ethnic group is still the largest major ethnic group in New Zealand, accounting for 76% of the population, according to the 2013 Census (Statistics New Zealand 2014). What the census figures also show is that among the most common languages spoken in New Zealand English is the dominant one, spoken by 96.1% of people who stated at least one language. Along with English, te reo Maori (the Maori language) and New Zealand Sign Language are further official languages of New Zealand. However, these languages

are spoken by only a small proportion of the population. In the 2013 Census, 148,395 people (3.7%) reported to speak te reo Maori and 20,235 people reported the ability to use New Zealand Sign Language.

Writing about regional variation in New Zealand English, Bauer (1994) notes that the surprising thing about it is "how little of it there is. Given the way in which New Zealand was settled from Britain, it might be expected that traces of Scottish, Irish and West Country dialect features (at least) would be found in different areas of New Zealand" (1994: 411). Some regional variation can be found in the South Island, in Otago and Southland, which were mainly settled by Scottish immigrants. But the linguistic traces they left are minimal. In this sense, New Zealand English can be seen as dialectally homogeneous, as Bauer suggests, "although there are social dialects of New Zealand English as there are of other varieties of English, and there may be at least differences of style between urban and rural speakers" (1994: 411; for stylistic variation in New Zealand English grammar see, for example, Hundt 1998).

Furthermore, New Zealanders believe that there is a distinct ethnic variety of Maori English (Bell 2000: 221; also see the quote from ICE-New Zealand at the very beginning of this section), but linguists have struggled to find clear and conclusive differences between Maori and Pakeha English since the 1960s. Numerous studies have addressed this question, but they found only quantitative and no qualitative differences (e.g. Benton 1991; Britain 1992; Meyerhoff 1994; Bell 1997, 2000; Schreier 2003). It is reported, for example, that Maori use more High Rising Terminal Contours than Pakeha (Britain 1992) and that they use more *eh* discourse tags (Meyerhoff 1994). Still, linguists describe the nature of Maori English as "[a]mong the most intriguing and elusive issues in the study of New Zealand English" (Bell 2000: 221) and that evidence of its existence is "at best tentative and ambiguous" (Benton 1991: 195). Some even deny its existence stating that "[t]here is no single identical variety of Maori English [...] It seems that it is more of a style of English than an actual separate dialect" (Gordon/Deverson 1998: 144-145). Schreier (2003) in a study of Maori English in the nineteenth century finds distinctive features which, he suggests, "originated as L2 learning processes in a language contact scenario [...] either through substratum effects, phonological transfer, or contact-induced adaptation" (2003: 388). But these features did not persist as Maori English converged with Pakeha English rather quickly from the nineteenth century onwards (ibid.).

2.2.3 Canada

> *Canadian English is a multi-ethnic language, spoken by people of every color and creed on earth.*
>
> (Boberg 2010: 25)

The earliest British contact with Canada was in 1497 when the Venetian mariner John Cabot discovered the eastern coast of Canada while exploring North America on behalf of King Henry VII. This event already sparked England's interest in establishing colonies in the Americas, but settlement began in earnest only later (Boberg 2010:

58).[5]

A number of small English settlements were established in a few regions beginning in 1610, namely in Nova Scotia, Newfoundland and around Hudson Bay. These first English-speaking settlers were mainly fishermen, fur traders, soldiers and some farmers with their families, but their number never rose much above one thousand inhabitants (Boberg 2010: 58/60). Large scale English-speaking migration began only in the eighteenth century.

In the seventeenth century, English-speaking people were not the only immigrants to settle in Canada, but there were also some French settlements. It was not long before French and British colonial interests began to conflict, culminating in the Seven Years' War (1756-1763). After the French surrender, the Treaty of Paris (1763) dictated France to cede to Britain its possessions in what is now Canada (Boberg 2010: 58). The British victory had important consequences: it not only sparked the first major wave of English-speaking migration to Canada but cut off further French immigration and initiated the long struggle of the French-speaking communities to survive and maintain their French identity and culture in an increasingly English-speaking continent (ibid.).

The first major wave of English-speaking migration began in the spring of 1783 when the first so-called 'United Empire Loyalists' left New York heading for Canada. The Loyalists are those American colonists who remained loyal to the British crown in the American Revolution (1765-1783) and emigrated during and immediately after it to what is today Canada. The majority of the Loyalists came from the middle and New England colonies, that is, Virginia, Maryland, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island and Massachusetts (Boberg 2010: 60-61). Most of them moved to Ontario (Dollinger 2012: 1861).

At the same time, direct immigration from Britain had also begun on a small scale, with most settlers coming from Scotland, particularly the Scottish Highlands. Note that these Highlanders were not necessarily English-speaking, but most of them spoke Gaelic (Boberg 2010: 65). The same is true for the many Irish people who emigrated to Newfoundland in the late eighteenth and early nineteenth century, the peak period of English and Irish emigration to the province. That is, like the Scottish Highlanders, many Irish spoke Gaelic when they arrived.

While Canada's English-speaking population was thus founded largely by Loyalist refugees in the eighteenth century, "its establishment as the dominant culture of modern Canada was assured in the nineteenth century by a more or less constant flow of direct immigration from Britain and Ireland" (Boberg 2010: 67). Rather than from political upheaval, these immigrants mainly fled from poverty, social dislocation and lack of economic opportunity, which was caused by a post-war recession setting in in 1815 (Boberg 2010: 68).

In the 1860s, British immigration declined, but it was to rise again towards the end of the century when Canada's vast western regions were opened up for settlement (cf. Dominion Land Act of 1872). Four major groups accomplished the settlement of these western provinces: migrants from Canada, especially Ontario, who were thor-

5 The present section presents only a very brief overview of the history of the English language in Canada. It is largely based on Boberg's (2010) monograph *The English language in Canada*, which I recommend to consult for more detailed information on Canadian English.

oughly Canadian by that time; immigrants from Britain, in particular from England; immigrants from other European countries; and further migrants from what is today the United States. Furthermore, the western land boom also attracted large numbers of Chinese immigrants, coming to Canada to help build the railroads (Boberg 2010: 88ff.).

The origins of Canadian English are difficult to exactly pin down. It would be necessary to find out where the immigrants precisely came from and which languages or dialects they spoke when they arrived in Canada. But record-keeping at the time was not easy and immigrant statistics are thus often not reliable or insufficient. Trying to account for the largely North American character of Canadian English, some researchers suggest that the Loyalists played a major role in establishing the basic pattern for Canadian English, thus recognizing the Founder Principle (Mufwene 2001, 2008), which suggests that the founder population in an ecology exerts strong influence on the shaping of the new variety (Bloomfield 1948; Chambers 1998; Dollinger 2008). Boberg agrees that the Loyalists indeed played a crucial role in what was to become the Canadian identity, but he also points to the fact that "comparatively little is known about exactly where the settlers came from or how they mixed with each other in new communities, much less how they spoke when they arrived" (2010: 100-101). While the Loyalists predominantly came from the middle and New England colonies, as mentioned above, this was "not to the exclusion of other regions" (ibid.). Furthermore, it may well be that many of the features which are today regarded as North American are actually directly derived from regional dialects of British English; that is, through the speech of the early British immigrants to Canada (Boberg 2010: 102).

The quotation from Boberg at the very beginning of this section describes the situation of Canadian English as it is perceived today, "a multi-ethnic language, spoken by people of every color and creed on earth" (Boberg 2010: 25). During the course of the twentieth and twenty-first centuries, Canada has become a multicultural society and Canadian English a multiethnic language. This is due to the fact that after the 1960s immigration came increasingly from non-traditional sources, in particular Asia, but also the Caribbean, the Middle East and Latin America (Boberg 2010: 97).

Canada's linguistic diversity is illustrated by the fact that more than 200 languages were reported as a home language or mother tongue in the 2011 Census of Population. These include the two official languages English and French, Aboriginal languages and immigrant languages. The latter are languages "whose presence in Canada is originally due to immigration" (Statistics Canada 2012: 1). In 2011, immigrant languages spoken as a mother tongue account for 19.8% of Canada's population (6.6 million people).[6] Among these, persons with an Asian language as their mother tongue make up the highest proportion, accounting for 56%. More than 40% of the immigrant-language population report a European language as their mother tongue. The top immigrant language reported in Canada is Punjabi, an Indo-Aryan language spoken in India, whose population amounts to about 460,000 persons. Romance languages other than French are also widespread, with Italian and Spanish reported as

6 In 2011, English is the mother tongue of 18.9 million people (or 56.9% of the population), while French is spoken as a mother tongue by 7 million people (21.3%). These data are based on single responses in the 2011 Census. Also see Table 2.2.

the mother tongue by more than 400,000 persons each. Within the Chinese language family, three main groups can be made out. Cantonese is the mother tongue of 389,000 persons and Mandarin that of 255,000 persons. Some immigrants simply reported Chinese as their mother tongue without specifying any further which dialect they spoke. This group comprises 441,000 persons (Statistics Canada 2012: 1-2).

Canadians who speak an immigrant language most often at home or as their mother tongue are predominantly found in Canada's metropolitan areas. Table 2.2 shows the number and proportion of Canadians with English, French or a non-official language as their mother tongue in Canada's ten largest metropolitan areas. The ten cities listed in Table 2.2 contain 55% of the national population (18.3 million people) but 81% of those people speaking an immigrant language as their mother tongue (5.3 million).[7]

Table 2.2: Population by mother tongue in Canada's ten largest metropolitan areas (single responses, 2011).

Metropolitan area	Total pop. N	English %	French %	Other %
Canada	**33,121,175**	**56.9**	**21.3**	**19.8**
Toronto, ON	5,541,880	53.8	1.1	41.8
Montreal, QC	3,785,915	11.6	63.3	22.0
Vancouver, BC	2,292,115	56.0	1.1	40.3
Ottawa, ON/QC	1,222,760	49.0	31.4	16.7
Calgary, AB	1,205,175	70.9	1.5	25.3
Edmonton, AB	1,146,600	74.5	2.2	21.3
Quebec, QC	756,400	1.4	94.9	2.9
Winnipeg, MB	721,120	72.2	3.8	21.4
Hamilton, ON	712,580	75.9	1.4	21.0
Kitchener, ON	472,090	74.5	1.2	22.6

Source: Statistics Canada, 2011 Census

As the numbers and proportions indicate, the linguistic make-up of Canada's ten largest cities is very different. The metropolitan areas of Toronto and Vancouver contain more than 40% of people who speak an immigrant language as their mother tongue. In Quebec, on the other hand, this group makes up only 2.9%. Quebec is an overwhelmingly francophone city, with 94.9% of its population having French as their mother tongue. This is mainly due to a programme of massive government intervention designed to prevent the gradual decline of French in the province. As a consequence, the English language has been in retreat since the new language laws came into effect. This development has been highly controversial.[8]

Ottawa, Canada's capital city, is located in Ontario, but its metropolitan area includes regions in Quebec. This bi-provincial nature of the city accounts for the rel-

7 Note that both percentages have increased since the 2006 Census. In 2006, the ten largest cities contained 53% of Canada's population and the immigrant population speaking a non-official language as their mother tongue accounted for 78% (Boberg 2010: 21).

8 See Boberg (2010: 6-19) for a detailed discussion of the language situation in Quebec.

atively large proportion of francophones (31.4%). In sum, it can be noted that immigrants to Canada seem to prefer English-speaking metropolitan areas over French-speaking ones as their destination.[9]

As a result of the large-scale immigration of people who speak languages other than English and French, new ethnic varieties of English have developed in Canada's metropolitan areas (cf. e.g. Boberg 2004; 2010 Chapter 5). The future will tell whether or in how far the immigrant languages will impact on the shape and development of Canadian English, thus reflecting "the multiethnic and multilingual character of urban Canada" (Boberg 2010: 105).

In addition to ethnic variation, there is also regional variation in Canadian English. It is commonly noted that Newfoundland English should be treated separately from mainland Canadian English. The variety of English spoken in Newfoundland is noticeably different because it was settled at a different time and by different groups of people. Traditional Newfoundland English is heavily influenced by southwestern English and southeastern Irish varieties and thus contrasts starkly with the North American speech of Ontario and western Canada (Boberg 2010: 26; see Clarke 2010 on Newfoundland and Labrador English).

As for standard Canadian English, it is generally considered as being largely homogeneous across mainland Canada (e.g. Bloomfield 1948: 63; Chambers 2006: 385; Dollinger 2012: 1860). However, recent research shows that there is some regional variation even within standard Canadian English (cf. Bolinger 2010, Chapters 4 and 5).

2.2.4 Singapore

> *I mean when you teach the school in England then how I mean like our English is not their their English ya.*
>
> (ICE-SIN:S1A-060)

The story of the English language in Singapore begins in 1819 when Singapore was acquired by the British East Indian Company and became part of the Straits Settlements (joining Penang and Malacca). The spread of English is strongly connected with the development of the education system because it was almost exclusively through the school system that the English language was acquired. In the early nineteenth century the first English-medium schools were established by private organizations, churches and charitable bodies (Gupta 1998: 110). But these schools were essentially for European and Eurasian children and the numbers of learners of English were rather low. It was only in the early twentieth century that the English language really started to spread, with Chinese children thronging the English-medium schools. Many teachers in Singaporean schools came from Malay and India and consequently there are still similarities in lexical and syntactic usage between Singapore English, Malaysian English and Indian English (Foley 1988: 4).

9 For a more detailed account of Canadian English in Canada's ten largest cities see Boberg (2010: 20-25).

In the early years, education was largely in the hands of private organizations, but already during the British period the government started to take over more and more control of the education system. This trend was continued after independence in 1965 and education is now under tight government control. Since 1987, all education under the government has been required to be in the medium of English (Gupta 1998: 115).

In Schneider's Dynamic Model (Schneider 2003, 2007), Singapore English has clearly gone through the process of structural nativization and has reached the phase of endonormative stabilization (phase 4). According to Schneider, this is more visible on the level of Singlish, the colloquial variety of English spoken in Singapore, but also in formal styles (2003: 265).

The Ethnologue lists 24 living languages for Singapore today and English is one of the four official languages, the others being Mandarin, Malay and Tamil. The importance the government ascribes to the language today can also be seen in the syllabus of English language teaching for primary and secondary schools of 2010. As guiding principles the Ministry of Education postulates (emphasis mine):

> Bilingualism is a cornerstone of our education system. Pupils learn both English and their own Mother Tongue language in school. English is the medium of instruction in our schools as well as a subject of study for all primary and secondary school pupils. English operates at many levels and plays many roles in Singapore. At the **local** level, it is the common language that facilitates bonding among the different ethnic and cultural groups. At the **global** level, English allows Singaporeans to participate in a knowledge-based economy where English is the lingua franca of the Internet, of science and technology and of world trade.
> Singapore's transformation into a knowledge-based economy, the rapid developments in technology, the **generational shift in home language** and an increasingly competitive international environment are some factors that make proficiency in English necessary for pupils. (Ministry of Education Singapore 2009: 6)

The syllabus notes two different functions that English has in Singapore - the global and the local. These two functions give rise to two different orientations and norms. The globalist orientation looks outwards and seeks homogeneity and similarity to other English varieties (Standard English). The localist perspective, on the other hand, seeks its norms inside and is rather separatist in nature (colloquial Singapore English or Singlish). Its major concern is for "a uniqueness of the English language in Singapore that must set its citizens apart from other English speakers" (Alsagoff 2010: 343). Depending on the context and on what they want to express, Singaporeans can choose between these English codes.

The syllabus, moreover, talks about 'the generational shift in home language'. This refers to the fact that the usage of English at home is becoming more prevalent, as the following census data illustrate.[10] Note that there is a parallel expansion of Mandarin, largely at the expense of the other Chinese dialects. English spoken as a home language is especially prevalent among the younger age groups (aged 5-14) and among university graduates (cf. Department of Statistics Singapore 2014; 2010 Census).

Because of Singapore's multicultural and multilingual make-up, early twentieth

10 The census data are from Leimgruber (2013: 3). For the Singapore Census of Population 2010 also see the website of the Department of Statistics Singapore at http://www.singstat.gov.sg/ (accessed: Sept. 2014).

century Singapore English has been restructured through contact with various other languages. In the beginning it was predominantly Bazaar Malay[11], Baba Malay[12], Hokkien and Cantonese that had the greatest impact on the shaping of Singapore English, today it is Mandarin.

Table 2.3: Language most frequently spoken at home in percentages.

	1980	1990	2000	2010
English	12	20	23	32
Mandarin	10	26	35	36
Chinese Dialects	60	37	24	14
Malay	14	13	14	12
Tamil	3	3	3	3

Interestingly, it is English rather than any of the vernaculars that is the "language for the construction and expression of the Singaporean (i.e. national) identity" (Lick/ Alsagoff 1998: 207). This is because English is the common language of the major ethnic groups in Singapore, that is, it brings citizens of different ethnic origins together and allows them to communicate and understand each other. Thus, the English language helps to form a Singaporean nation rather than a Malay, Indian or Chinese nation.

2.2.5 The Philippines

> *English is the common language to use in in the class but uh if you're talking with fellow uh Filipinos then you can converse in Tagalog.*
>
> (ICE-PHI:S1A-091)

The historical origins of Philippine English can be dated to 1898 when the United States started to occupy and colonize the Philippines. American teachers were sent to the country at the beginning of the twentieth century to teach the Filipinos the English language. They had an important impact not only as teachers but also as teacher-trainers and by 1921 91% of all teachers were native-born Filipinos (Bolton/ Bautista 2008: 4). The English language spread rapidly in the Philippines, a spread which "was unprecedented in colonial history, for within the space of 41 years, the American regime had done more to spread English than the Spanish Government did in 333 years (1565-1898) of colonization, for at the end of the Spanish Period, only 2% spoke Spanish" (Gonzalez 1997: 28).

11 Bazaar Malay is a pidginized form of Malay and the main *lingua franca* in the early twentieth century (cf. Lim/Foley 2004).

12 Baba Malay is a Malay-based creole spoken by Straits-born Chinese, about 500 years old but now nearly extinct (cf. Lee et al. 2009)

When the English language arrived in the Philippines there was no national language but over 100 indigenous languages spoken by Filipinos. In 1937 - after many years of debate - Tagalog, the language of one of the most important indigenous ethnic groups (next to the Cebuanos and Ilocanos), was chosen as the basis for the national language, Filipino. Tagalog and Filipino basically differ only with respect to the lexicon, with "the Filipino lexicon being systematically expanded by the Institute of National Language" (Himmelmann 2005: 350). Although most people in the Philippines realize that Filipino is in fact "Tagalog with extras", they have come to accept it as their national language (Kirkpatrick 2012: 22).

After independence in 1946, English was retained as official language in government and education but was increasingly used alongside the national language. The linguistic repertoire of educated Filipinos was basically dominated by the English language up until the 1970s, but then national fervour became so strong that the domains of the English language were reduced in favour of Filipino. The domains it is still used in today include higher education, the print media (21 out of 28 daily newspapers are in English), business transactions in internationally-oriented companies, diplomacy and international relations (Gonzalez 2008: 22). For informal communications, Filipinos do normally not use the English language but rather the national language Filipino or their home vernacular (also see the quote from ICE-Philippines at the very beginning of this section). When they have to speak English in an informal context, they often code-switch, as the nearly 2000 examples of indigenous inserts in the 'private dialogues' files of ICE-Philippines impressively document. Here is an example.

(2.1) B: What happened to Brother Andrew?

A: *Oo nga e* [laughter]

B: *Ano'ng nangyari kay* Brother Andrew.

A: *Wala pang balita e. Saan kaya siya.* I have to go to the next room for my meeting.

B: Don't worry.

A: *Hay 'naku* Sir [laughter] You don't know people in La Salle.

B: No what is this?

A: They are very punctual.

(ICE-PHI:S1A-003)[13]

13 I am indebted to Ariane Macalinga Borlongan for translating the Tagalog clauses in this dialogue for me. The translations are as follows:

B: What happened to Brother Andrew?

A: *Yes, truly* [laughter]

B: *What happened to* Brother Andrew?

A: *There is no news yet. Where can he be*? I have to go to the next room for my meeting.

B: Don't worry.

A: *Gosh*, Sir [laughter] You don't know people in La Salle.

Code-switching and code-mixing among English-knowing bilinguals is in fact so widespread that the resulting code has been given its own name, 'Taglish', a mixture of English and Tagalog which often tends to be used as the unmarked code of choice in informal conversation, for example, in Manila (Bolton 2003: 201).[14]

As for Schneider's Dynamic Model, the position of Philippine English is difficult to precisely specify and different views are voiced in the literature. While Schneider (2007) claims that Philippine English is in phase 3, probably approaching phase 4, Borlongan (2011) argues that this evolutionary phase is already well under way. Collins and his colleagues (2014), in a very recent study, find evidence of both exonormative and endonormative orientation in Philippine English modals and quasi-modals. They conclude that their results reflect "the somewhat ambivalent evolutionary status of PhilE, with opinions divided on the issue of whether its entry into phase 4 ('endonormative stabilization') of Schneider's (2007) evolutionary scale is merely incipient or well established" (2014: 85). Martin (2014) also argues that among Philippine English speakers "[s]trong preferences for its American parent, as well as the culture that comes with American English remain" (2014: 81). While acknowledging that Philippine English "seems to have found its place" among the educated class, she is doubtful about how widespread the acceptability of the English variety really is (2014: 79). The language remains associated with the educated class and is not an identity carrier for most Filipinos (as is the case with Singapore English). Martin concludes that Philippine English has indeed developed into a nativized form, but "[w]hether or not that English progresses into a variety of Endonormative Stabilization remains to be seen" (2014: 81).

Kirkpatrick (2012) presents a more complex picture of Philippine English within Schneider's model. He suggests that, linguistically, Philippine English shows evidence of the final stage of differentiation. He argues that there is a continuum of English varieties ranging from informal Taglish to a more formal, educated variety; and educated Philippine English speakers may use different dialects from this continuum depending on the communicative situation. Sociolinguistically, however, Philippine English is somewhere between stage two and stage three, Kirkpatrick claims, because the idealized classroom model is still American English rather than a local variety (2012: 17).

According to the Ethnologue, there are 181 living languages in the Philippines today. The national language Filipino is not the L1 for all Filipinos but has 45 million L2 speakers while English is spoken as a L2 by 40 million people in the Philippines.[15] Typical Filipinos are minimally bilingual, more often even trilingual, if Filipino is not their mother tongue. At home, they speak their mother tongue, which may be a minority language; for intranational communication they use Filipino; and for international relations or in school they use English.

The issue of the English language in relation to the national language has heavily been debated in past decades. Filipino intellectuals continue to voice concerns about

14 For a detailed discussion of the interrelationship between English and Tagalog in the Philippines also see Thompson (2003). He examines the nature of and motivations for code-switching among Filipinos and in the mass media.

15 See the entry "Philippines" on the Ethnologue's homepage at http://www.ethnologue.com/country/PH (last accessed: Aug. 2014).

the power and prestige of English and claim that it contributes to sustain social and economic inequality and that it hinders the creation of an authentic sense of nationalism. But there are also people who realize that knowing the English language and improving their proficiency may be of help in the job market, that is, they are aware of the need of the L2. This development has resulted in determined attempts to attain the quality of English.

It has been noted before that from early on the Filipinos learned the English language from Filipino teachers. Hence, the source language American English was never completely replicated at any time, but from the beginning there were many local varieties of Philippine English based on the L1 of the speaker (Gonzalez 2008: 20). Philippine English is now well-recognized as an autonomous variety of English, but it still faces problems of legitimation, that is, the issue of standardization has yet to be resolved. According to Gonzalez (2008: 21), international communications, the mass media and the print medium "may promote the most acceptable variety" of Philippine English, which may then become the standard.

2.2.6 Jamaica

> *A: But Jamaican English why is Jamaican English so peculiarly interesting then? Jamaican English is no different to*
> *B: It is very different <,> my man*
>
> (ICE-JAM:S1A-091)

In 1655, the British attacked Jamaica taking over the island from the Spanish, who, vastly outnumbered by the invaders, could mount little resistance. The white population that settled in Jamaica after the British conquest was drawn mainly from Barbados, the Leeward Islands[16], Suriname and England (Lalla/D'Costa 1990: 14ff.; Holm 1994: 341; Rosenfelder 2009: 11). A large number of the originally 1,600 settlers died within a year, but due to further immigration Jamaica's population rose again and by 1658 consisted of 4,500 whites and 1,400 blacks (Holm 1994: 341). This ratio was to shift in the coming decades due to massive importation of new slaves. At that time, sugar was the main crop in Jamaica and large plantations were established wherever possible. Yet, the cultivation of sugar was so labour-intensive that more slaves were needed. By 1739, the ratio of white to black had shifted to about 1:12 (Lalla/D'Costa 1990: 22).

Some creolists argue that these circumstances - the large slave population and the sugar plantations with their rigid hierarchical system - provided the grounds for the establishment of a creole, a "distinct language system with words from English but with phonology, semantics and morphosyntax influenced by African languages and other forces" (Holm 1994: 328). The slaves transported to Jamaica came from a number of different ethnic groups, particularly from West and Central Africa, and thus had no common language. Furthermore, they received only little input from En-

16 The Leeward Islands make up the northern part of the great arc of the Lesser Antilles, consisting of the islands of St. Kitts, Nevis, Barbuda, Antigua, Montserrat and Anguilla.

glish. Hence, in order to be able to communicate with the British colonizers, the white indentured servants and one another they created a simplified version of English (Alleyne 1984; Holm 1989: 470). Another theory of Jamaican Creole genesis, on the other hand, holds that the creole has essentially been transplanted from Barbados, that is, it emerged as the continuation of a pre-existing creole (McWhorter 2006: 110). McWhorter, in fact, claims that all Atlantic English-based creoles can be traced to this common ancestor (ibid.).

During the eighteenth and nineteenth centuries, Jamaica "became increasingly creolized and distanced from the African past" (Lalla/D'Costa 1990: 31). The creole population was slowly growing and eventually outnumbering the African population (Lalla/D'Costa 1990: 25-26). This development was accompanied by a growing importance of Jamaican Creole as a means of communication and it "gradually became clear that somehow foreigners' speech ('broken English', for example) had taken root and become the local language of blacks, influencing the speech of local whites as well" (Holm 1988: 17). The end of British slave trade in 1809-1810, then, marked the beginning of the decline of direct African influence in Jamaica; the number of slaves declined and that of freed blacks and "people of color" rose, important changes that reflect "the movement toward a creole society that accepted Jamaica as its home" (Lalla/D'Costa 1990: 26).

After the abolition of slavery in 1834[17] many slaves moved away from the plantations and settled on the island wherever they could find land to buy, establishing free peasant villages in which more conservative forms of Jamaican Creole came to be preserved (Senior 2003: 199-200; Cassidy/Le Page 1980: xlii; Patrick 2007: 127).

During the nineteenth century Jamaican Creole became increasingly influenced by English, which was mainly due to Christianization and the establishment of schools. The white missionaries sent to Jamaica worked closely with the slaves and the free blacks and browns, who thus came into direct contact with speakers of middle-class varieties of British English, especially varieties from the north and midlands. The speech of these missionaries provided the model for many free blacks and browns who hoped to gain social advancement through church and school. Thus, high speech and formal utterances were influenced by biblical language and prayer-book language (Lalla/D'Costa 1990: 29-30).

At the same time, the establishment of new schools also led to growing influence of (British) English on Jamaican Creole. As a result, literacy rates doubled from around 30% to approximately 60% of the Jamaican population (Senior 2003: 173). And since the curricula were British-based and the teaching staff imported from Britain, English became "the model language towards which the Creole moved" (Cassidy/Le Page 1980: xlii; see also Senior 2003: 174).

The close and continued contact between Jamaican Creole and the English language resulted in a continuum of varieties, with Jamaican Creole (the basilect) and standard Jamaican English (the acrolect) making up the two poles of the continuum. Both poles are rather idealized abstractions, "a collection of features most like stan-

17 The Emancipation Act was actually already passed on 29 August 1833, granting that all slaves be free from 1 August 1834 "but requiring them to undergo a period of 'apprenticeship' to their former masters (who were compensated monetarily for their loss) for another four years" (Rosenfelder 2009: 16).

dard Englishes (the acrolect) or most distant from them (basilect)" (Patrick 2004: 408). In between these poles lies a number of further varieties, also called the mesolects, which constitute "the continuum of everyday speech: a series of minimally differentiated grammars with extensive variation", as Patrick puts it (2004: 408-409).

The linguistic variability found in Jamaica was first described in terms of a continuum by DeCamp (1971), writing that[18]

> [t]he varieties of Jamaican English themselves differ to the point of unintelligibility; but some Jamaican English is mutually intelligible with standard English. [...] Further, in Jamaica there is no sharp cleavage between Creole and standard. Rather there is a linguistic continuum, a continuous spectrum of speech varieties ranging from the 'bush talk' or 'broken language' of Quashie to the educated standard of Philip Sherlock and Norman Manley. [...] Each Jamaican speaker commands a span of this continuum, the breadth of the span depending on the breadth of his social contacts [...]. (DeCamp 1971: 350)

The creole continuum model became generally accepted by creolists and is by now regarded as a valuable tool for the description of the extreme variability found in Jamaican speech (and other creoles; also see Bickerton 1973, 1975 for an early proponent of the model). Yet, there are also creolists who challenge the idea of the continuum, suggesting to see Jamaican Creole and Jamaican English as two discrete and self-consistent grammars (Bailey 1971; Lawton 1980; Devonish 1998, 2003).[19]

While the creole continuum model was treated as a purely linguistic phenomenon by earlier studies, more recent research integrates social factors and regards the creole continuum as a sociolinguistic model (Sand 1999; Patrick 1999, 2004; Deuber 2014):

> Social stratification in Jamaica is crucial to understanding the extreme variability of contemporary Jamaican speech. The complex linguistic situation can be related to an equally intricate web of social relations, using the model of the creole continuum. (Patrick 2004: 408)

A similar position is taken by Deuber who suggests that "spoken English in the Caribbean is characterized by considerable variability and that this may be seen as embedded in a continuum of synchronic sociolinguistic variation" (2014: 11). These studies also show that mesolectal and basilectal forms are often used in speech dominated by acrolectal forms for stylistic effects. Somewhat reinterpreting the creole continuum, Deuber proposes that the notion of the continuum is appropriate to describe the range of varieties found in Jamaican speech only if social and stylistic connotations are taken into account. Acknowledging that English and Jamaican Creole necessarily share forms and that there is a partial overlap of the two grammars, she concludes:

---

18 DeCamp (1971) actually uses the term 'post-creole speech continuum' to describe the language situation in Jamaica. Assuming that pidgins, creoles and the creole continuum represent different stages of a life cycle, he proposes four alternative solutions for the final stage: (1) the creole "can continue indefinitely without substantial change"; (2) it "may become extinct"; (3) it "may evolve into a 'normal' language"; and (4) it "may merge with the corresponding standard language" (DeCamp 1971: 351). Jamaica represents the last alternative, according to DeCamp, and in order to distinguish it from the first alternative this stage of development is called 'post-creole'. Bickerton (1973) questions DeCamp's model writing that "since something marginally, if at all, different from the original creole language frequently constitutes the basilect of the continuum, 'post-' can be misleading for Jamaica" (1973: 640). See Sand (1999: 50ff.) for a more detailed discussion of these different viewpoints.

19 See Sand (1999) for a more detailed discussion of the two different perspectives.

> Apart from that, however, the perspective adopted in the present study supports the view that English and Creole can be separated as linguistic systems and that it is in the spectrum of social and stylistic variation linking the extreme varieties that the nature of the Creole continuum lies. (Deuber 2014: 242)

Deuber hastens to clarify that this reinterpretation of the creole continuum does not only result from a shift in the author's perspective but also "responds to the way speakers have been reinterpreting and adapting the range of linguistic variation characteristic of Caribbean Creole continua" (ibid.).

A different situation characterizes the relationship between English and creole in writing and computer-mediated communication (CMC; e.g. emails and internet discussion forums and blogs). Mair (2002b), for example, observes that the role of the creole is very limited in written Jamaican English. If it is present, it is clearly marked off from the dominant English text by quotation marks or other metalinguistic clues, indicating that it is not the writer's own words; or it occurs in cartoons or proverbs (Mair 2002b: 36). Mair furthermore suggests that, contrary to speech, a diglossic situation characterizes the relationship between creole and English in writing:[20] "the concept of diglossia [...] which has proved insufficient to describe the complexities of spoken usage serves quite well to describe written practice" (ibid.). In CMC, on the other hand, creole-influenced writing is more prevalent. While Mair (2002b: 56) notes a "continuum-like writing practice" in such texts, Hinrichs (2006) questions the usefulness of the continuum and proposes a code-switching analysis instead:

> the adaptation of mesolectal and basilectal forms for use in CMC has generally not preserved the ordered and small transitions of the spoken continuum, but replaced them with two separate codes which are in principle easy to distinguish, even if occasional difficulties or overlaps are encountered in individual lects. (Hinrichs 2006: 40)

In sum, present-day research yields more fain-grained results with respect to the relationship between English and creole. It shows that social and stylistic factors as well as register variation need to be taken into account to arrive at conclusive descriptions of speech and writing in Jamaica.

Another important issue that needs to be considered in the context of language in Jamaica is the question of prestige. Jamaican Creole has commonly been held in little esteem, described as 'bad English', an imperfect variety that needed to be corrected through education. British English provided the model and was considered superior to native Jamaican languages, especially prior to political independence, which Jamaica gained in 1962. Differences between standard and non-standard English as well as creole features were of no interest. In fact, the creole was disregarded as inherent in Jamaican society and as a language in its own right (Shields 1989). The present-day situation in Jamaica shows that such prejudice against the creole still persists as letters to the editor or newspaper columns, for example, reveal. Yet, there is also evidence that attitudes have been considerably transformed over the past few

20 The term 'diglossia' (Greek *di* = 'two', *glossa* = 'language') as it is used in sociolinguistics goes back to Ferguson (1959). It describes a "relatively stable language situation" (Ferguson 1959: 336) with a high (H) and a low (L) code being strictly separated according to functional domain. H is the code used in formal, written, official, ceremonial, solemn, institutional and legal domains. L is everyday language, spoken in family and other intimate and informal settings.

decades (Deuber 2014: 30; also see Beckford Wassink 1999). The positive attitude towards Jamaican Creole, a symbol of national identity, is reflected in the fact that it has moved into more domains recently, domains that had formerly been reserved for standard English only (Sand 1999: 73; Mair 2002b: 32). It is today used in schools and in the media, for example, on the radio by DJs, in public service messages, interviews, phone-ins; or in newspapers in cartoons, local gossip columns and direct speech (Sand 1999: 73). What makes the situation of Jamaican Creole difficult, however, despite its by now manifold functions, is the lack of standardization and homogenization. Outside linguistics, there is no common orthography, but people rely on eye-dialect for written purposes (Sand 1999: 75; Mair 2002b: 33). Furthermore, it was only in the twenty-first century that the Jamaican government has seriously begun to address the question of discrimination on the ground of language and "to explore language planning and recognition of Jamaican Creole as a national language" (Patrick 2004: 408).

Two final points are worth mentioning with respect to the situation of English in Jamaica. First, note that Jamaican Creole was influenced by a number of non-standard regional English varieties (e.g. the working-class speech of London, Bristol, the West Midlands and Liverpool, Scots and Irish English) with many dialect features surviving in Jamaican Creole (Patrick 2007: 127). And second, it is important to note that the influence of American rather than British English grew in the Caribbean area at the beginning of the twentieth century as a result of the United States' emergence as a world power (Holm 1994: 354). Today, American English is probably the most significant prestige dialect in the region (Mair 2002b: 34).

2.2.7 India

> *Or in other words <,> uh if you don't know English <,> if you want to communicate with people from other cities you need to know around uh ten fifteen languages <,> Whereas if you know English you could serve <,> I mean it could serve your <,> purpose throughout India and even outside the world.*
>
> (ICE-IND:S1A-025)

Indians got first into contact with the English language in 1579 when the Jesuit missionary Thomas Stephans arrived at the subcontinent. The spread of English really began in 1600 with the establishment of the East India Company. British merchants were granted a royal charter to trade with India, which resulted in the gradual establishment of trading posts all over the subcontinent and the introduction of English as a means of communication. Particularly important agents of the introduction of English were the missionaries who established the first English-medium schools. However, contact between the settlers and the local people was rare during the first one hundred and fifty years of British involvement. The settlers considered themselves genuinely British people and the local people viewed English as a foreign language (Sedlatschek 2009; Mukherjee 2007).

The position of English began to change in the eighteenth century when Britain got more and more political control over the Indian subcontinent. As a result, English was firmly present in the domains of administration and bureaucracy. In the nineteenth century, the presence of English increased as it became established in further important domains, such as commerce, the print media, academic and literary writing and education (Sedlatschek 2009: 11). It was especially the use of English in education and bureaucracy that became crucial and marked "the beginning of Indians' English" (Krishnaswamy/Burde 1998: 89).

In the beginning, all educational activities were taken up by the missionaries, but in 1813 education was brought directly under the control of the East India Company. This led to the first major language debate in India termed the 'Anglicist-Orientalist' debate (Sailaja 2009: 103). This debate was ended by Thomas Babington Macauly's "Minute on Indian Education" presented in 1835. At that time, Macauly was a member of the Supreme Council of India. The "Minute" was addressed especially to those Council members who believed that Indian students should continue to be educated in Sanskrit and Arabic as well as English.

> How, then, stands the case? We have to educate a people who cannot at present be educated by means of their mother-tongue. We must teach them some foreign language. The claims of our own language it is hardly necessary to recapitulate. It stands preeminent even among the languages of the West. [...] Whoever knows that language, has ready access to all the vast intellectual wealth, which all the wisest nations of the earth have created and hoarded in the course of ninety generations. [...] Whether we look at the intrinsic value of our literature or at the particular situation of this country, we shall see the strongest reason to think that, of all foreign tongues, the English tongue is that which would be the most useful to our native subjects. (Greenblatt et al. 2006: 1610)

Macauly's "Minute" was accepted by the government and English became the language of higher education, which continues to be the case even today.[21] With the firm establishment of English in the education system, the position of the language changed, as Sedlatschek states: the "'Minute' finally shifted the status of English from being a foreign language to being an official language" (2009: 13).

Interestingly, the National Congress, which was formed in 1885 and which aimed at leading India into independence, used the English language against the rulers themselves. That is, English served as a link language for the Indian people, a function which is still recognized as important, as the quote from ICE-India at the beginning of this section shows. After independence in 1947, there were fierce debates among the national leaders on which language should be the national language. Eventually, Hindi was adopted as the official language with English as an associate official language. Ironically, the constitution of the new republic was written in English, but the language is not part of the 22 scheduled languages in the constitution (cf. Government of India, 2001 Census) and a Hindi translation of the constitution was provided only quite some time later.

Due to its history as the dominant language of administration, education and the media, English has emerged as the language of the highly influential classes of Indian

21 The present situation is different in primary and secondary schools, where English is usually no longer the medium of instruction (Mehrotra 1998: 7).

society. As a result, English has often been associated with power and success, and, interestingly, in spite of being the language of the former colonizers, still enjoys a high prestige in India and is valued as an international language. Additionally, it has also made its way into more private domains and is today also used as a link language in the domains of family, friendship and neighbourhood (Sedlatschek 2009: 22).

As for Schneider's Dynamic Model, Indian English is seen as "an example of an evolutionary steady state in phase 4 with some coexisting features of phase 3" by Mukherjee (2007: 163). Schneider, on the other hand, is more reluctant in reliably identifying features of phase 4: "A few factors are foreshadowing endonormative stabilization, but they are disputable or weak; they should therefore not be overestimated" (2007: 171).

According to the Ethnologue, there are 447 living languages in India.[22] The language families that are present in India include Indo-European (in particular Indo-Aryan; 77% of the population), Dravidian (21%), Tibeto-Burman and Austro-Asiatic (1% each). Hindi is the most widespread language in India with 422,048,642 mother-tongue speakers (2001 Census); English is the mother tongue of 226,449 Indians, a very small proportion considering that the overall population in India amounts to over 1.2 billion people. The number of L2 English users, however, is quite high with around 200 million speakers (Crystal 2003), outnumbering English speakers in Australia and New Zealand and making India the country with the largest number of non-native speakers in the world (Mehrotra 1998: 1).

2.2.8 Hong Kong

> *Yeah uh in for me Hong Kong English means uhm English used in Hong Kong by Hong Kong people. But uhm I think uhm we seldom use English in our daily life. We seldom talk in English it's rather odd to speak in English uhm during our daily life.*
>
> (ICE-HK:S1A-037)

Hong Kong has a history of linguistic contact with English that dates back to the early seventeenth century when the first British trading ships reached Macao and Canton. A distinct variety of pidgin, also called 'Chinese pidgin English', developed and was used for commercial purposes between Europeans and Chinese traders, merchants and shopkeepers (Bolton 2003: 157). The acquisition and use of English shifted from a pidgin to a 'standard' variety after Britain had taken possession of Hong Kong in 1841 and missionary schools were established, in which both Chinese and 'western' subjects were taught. The system of government and missionary schools, which slowly developed over the second half of the nineteenth century, created 'elitist bilingualism' because the schools mainly catered for Chinese children from elitist background. This was to change in the 1970s when government reforms established a system of education that gave every child the opportunity to gain an education and learn at least

22 See the Ethnologue's homepage at http://www.ethnologue.com/country/IN (accessed: Aug. 2014).

some English. As a result, 'elitist bilingualism' was replaced by 'mass bilingualism' (or 'folk bilingualism') (Bolton 2003: 199).

Before 1998, instructions in primary schools were mainly through the medium of Cantonese. In secondary schools and universities, on the other hand, the medium of instruction was mainly English. But in schools which claimed to be 'English-medium' the prevailing classroom language was in reality often a 'mixed code' (i.e. a mix of Cantonese and English) and the quality of learning in these schools caused disquiet (Graddol 2013: 33). In 1998, the government decreed that only those schools which could demonstrate that they had enough competence would be allowed to teach in English and consequently the majority of secondary schools became Cantonese-medium. A similar situation holds for Hong Kong's universities, where "[t]here may be a mismatch between students' English proficiency and the aspiration of Hong Kong universities to be world class English-medium institutions" (Graddol 2013: 37). Note, furthermore, that informal use of English outside the classroom and lecture hall is low, as the quote from ICE-Hong Kong cited at the beginning of this section also hints at.

During the period of British colonial rule (1842-1997), English had the status of the official language of the government and law. It was only in 1974 that Chinese was also recognized as an official language, but it soon became more dominant in official domains, especially in the years prior to the 1997 handover of Hong Kong to the People's Republic of China. Its strengthened position also becomes obvious in the constitution of the Hong Kong Special Administrative Region adopted in 1990. Article 9 states that "[i]n addition to the Chinese language, English may also be used as an official language by the executive authorities, legislature and judiciary of the Hong Kong Special Administrative Region" (The Government of the Hong Kong Special Administrative Region 1990: 7).

In Hong Kong, the dominant spoken Chinese language is Cantonese, and increasingly so, as the census data of the years 1991, 2001 and 2011 illustrate (cf. Table 2.4).[23] Yet, officially Hong Kong is 'biliterate and trilingual'. The term 'trilingual' refers to the fact that next to Cantonese and English, which function as co-official languages, Putonghua has increasing currency in Hong Kong. Putonghua is spoken Mandarin, the official spoken language of mainland China. The term 'biliterate' includes English and written Chinese. Written Chinese in Hong Kong differs from mainland China in that the former still mainly uses 'traditional letters', while the latter has introduced 'simplified letters'.[24]

The numbers of speakers of English and Putonghua are also rising, which may be due to several reasons. First, the government and the public of 'cosmopolitan' Hong Kong have realized the importance of English as a global language. Putonghua, on the other hand, has been promoted for many years by the Beijing government as the 'national language' and it is the usual medium of education in mainland China. Note,

23 The figures of the years 1991 and 2001 are from the "2001 population census: summary results", and the figures of the year 2011 are from the "2011 population census: summary results" (cf. The Government of the Hong Kong Special Administrative Region, Census and Statistics Department 2001 and 2012).

24 Note, however, that the Hong Kong government provides three versions of its website in three languages: English, 'simplified Chinese' and 'traditional Chinese'.

however, that these two languages are perceived by Hong Kongers to have instrumental value only, while they have an emotional attachment to Cantonese (Gisborne 2009: 152).

Table 2.4: Hong Kong population aged 5 and over by usual language, 1991, 2001 and 2011.*

	1991		2001		2011	
	N	%	N	%	N	%
Cantonese	4,583,322	88.7	5,726,972	89.2	6,095,213	89.5
Putonghua	57,577	1.1	55,410	0.9	94,399	1.4
oth. dialects	364,694	7.0	352,562	5.5	273,345	4.0
English	114,084	2.2	203,598	3.2	238,288	3.5
others	49,232	1.0	79,197	1.2	106,788	1.6
total	**5,168,909**		**6,417,739**		**6,808,433**	

* The figures exclude mute persons.

Cantonese is essentially a spoken language and for educated Cantonese speakers standard written Chinese is the written form they use in most contexts. Written Cantonese may be used, however, in more informal situations, for example, among friends. The increasing use of Cantonese in Hong Kong - even in its written form - may also be seen as a reaction towards the People's Republic of China's promotion of Putonghua as the national language and of simplified letters in writing. Hong Kongers see this as a threat to their cultural identity. In a very recent study on the language landscape of Hong Kong (i.e. the language of official signs, advertisement and announcements in, for example, underground trains and lifts), Danielewicz-Betz and Graddol (2014) have found that Cantonese, traditional letters and English play an important role for Hong Kongers to demarcate their identity from that of mainland Chinese.

The question of whether there exists an autonomous (or semi-autonomous) variety of Hong Kong English - in the way that other Asian Englishes such as Indian English, Singapore English and Philippine English are recognized - has been the subject of ongoing debate. Tay (1991: 327), for example, states that "[t]here is no social motivation for the indigenisation of English in Hong Kong". In a similar way Johnson (1994: 182) notes that "[a] Hong Kong variety of English has been mentioned in the international literature [...] and in Hong Kong itself" but that this notion has received little support. He adds that "[t]here is no social or cultural role for English to play among Hong Kong Chinese; it only has a role in their relations with expatriates and the outside world" (ibid.). On the other hand, there are also academics who recognize Hong Kong English as an autonomous variety, yet an 'emerging' or 'developing' one (e.g. McArthur 1987; Bolton 2000, 2003; Kirkpatrick/Xu 2002; Gisborne 2009). Its standards tend to be exonormative but there are also distinct "Hong Kongisms" in everyday speech, which may be a sign of an emerging local norm. While Schneider (2003) suggests that the variety is on its way to structural nativization, others are more reluctant and state that it has yet to be seen when or even whether Hong Kong English will be moving towards the stage of endonormative stabilization (Groves 2009).

The main language English comes into contact with in Hong Kong is Cantonese. But due to large scale migration from mainland China in the 1990s there are also some other Sinitic languages spoken (e.g. Puthongua, Hokkien). Additionally, there is a large group of Filipino domestic workers, who bring not only their variety of English but also Austronesian languages into Hong Kong.[25]

25 According to the Asian Migrant Centre, there are more than 173,000 Filipino domestic workers in Hong Kong in 2015 (http://www.asianmigrantcentre.org/; accessed: Nov. 2015).

CHAPTER 3

Information packaging

> *I have been using the term packaging to refer to the kind of phenomena at issue here, with the idea that they have to do primarily with how the message is sent and only secondarily with the message itself, just as the packaging of toothpaste can affect sales in partial independence of the quality of the toothpaste inside.*
>
> (Chafe 1976: 28)

Speakers organize their utterances as components of a discourse, that is, they specify a structural unit and organize it in such a way as to relate it to the preceding discourse, thus achieving coherence. One way of achieving coherence in discourse is by drawing informational links between what is being said and what has been said before. Through these links the processing of the utterance becomes easier for the addressee and he can establish relationships between entities of the current conversation and the preceding discourse. For example, the use of the definite article marks a noun phrase as familiar and thus serves as processing signal to the addressee. The same is true for anaphoric pronouns, which pick up an entity that is already known and thus draw a link to the preceding discourse. Furthermore, speakers may want to draw the addressee's attention to a certain unit in the sentence. They can achieve this by using prosodic or syntactic devices, for example, by placing a higher pitch on the element they want to emphasize or by placing it in a position in the sentence where it would normally not occur.

The terms information structure or information packaging have come to be used as cover terms for the field of linguistic research that investigates the phenomena outlined above. The former term goes back to Halliday (1967) and his seminal paper on information structure. The latter term has been introduced by Chafe (1976) who used the term 'packaging' to refer to linguistic phenomena that "have to do primarily with how the message is sent and only secondarily with the message itself, just as the packaging of toothpaste can affect sales in partial independence of the quality of the toothpaste inside" (1976: 28; cf. the quote at the very beginning of this section).

This chapter consists of four parts. The first part will outline the theoretical concepts and technical terms developed for the study of information structure. The second part will introduce the non-canonical syntactic structures which speakers use to mark the information status of the constituents and which are the subject of the present study. This will be followed by a review of literature on the constructions at

issue. The final part of this chapter will give an overview of information-packaging strategies in some of the major background languages.

3.1 Basic concepts

3.1.1 Word order

In English, basic sentences have the word order subject-predicate with the predicate consisting of an object, complement or obligatory adverbial. The former case is illustrated in sentence (3.1a).[1] The proposition of sentence (3.1a) can be expressed in different ways, for example, by re-arranging the word order and placing the object in initial position, as in sentence (3.1b).

(3.1) a. Tom loves apples.

b. Apples Tom loves.

These two sentences express the same propositional content, but they differ in the way the information is organized in the sentence. The structure of sentence (3.1a) suggests that the agent *Tom* is known, while the newly added information about him is that he loves apples. The initial placement of *apples* in sentence (3.1b) indicates that, here, the entity *apples* is known (*What about apples? Apples Tom loves.*), while the fact that Tom loves them is the new contribution. Furthermore, the two sentences differ in what or who they are about, that is, they differ with respect to the topic. Sentence (3.1a) says something about Tom and so he is the topic of the sentence (assuming an intonational contour that is normal for basic declarative sentences, i.e. prosodic prominence is given to the final phrase) while the initial placement of *apples* in (3.1b) suggests that the sentence is about apples.[2]

The two sentences cannot be used interchangeably, but their use and interpretation depends on the current discourse. Speakers make a choice depending on what part of the message they want to emphasize or on what they believe to be known to the addressee, that is, the speaker knows that there are several factors in a clause which he "must manipulate as he speaks, so as to be able to get his message across with due consideration to the current state of his listener's mind" (Chafe 1976: 55). The clause-initial placement of the object *apples* in (3.1b), for example, works felicitously only in contexts where *apples* is contrasted with some other entity, which has previously been mentioned in the discourse, or if the term *apples* itself has been mentioned before; for example, in a question like *Does Tom love apples?* or in a statement like *Tom hates apples.* Sentence (3.1b) is thus unlikely to occur in discourse initial contexts, whereas this may well be the case for sentence (3.1a).

1 The following discussion is largely based on Vallduví and Engdahl (1996: 460f.)

2 As will be shown below, the fronted constituent *apples* in (3.1a) can also be the focus, being contrasted with some other entities (*What kind of fruit does Tom love? APPLES he loves.*).

3.1.2 Given - new

As has been illustrated in the previous section, an important assumption underlying discussions on information packaging is that a sentence typically consists of two parts, one containing old or given information and the other new information. Many languages tend to structure sentences according to the principle that given comes before new information, that is, information that the speaker thinks is known to the addressee is placed before the new information in the sentence (e.g. Halliday 1967: 205; Erteschik-Shir 2007: 7). Research on the correspondence between sentence position and givenness goes back to the early Prague School work on syntax and discourse function (Ward/Birner 2004: 155). Since then much research has been done on givenness, but there is a lot of disagreement and confusion about a precise definition of the concept. For Chafe (1976), for example, the notion of consciousness plays a crucial role in distinguishing between given and new information. Given or old information "is that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance" while "new information is what the speaker assumes he is introducing into the addressee's consciousness by what he says" (Chafe 1976: 30). According to this understanding, the term 'new' refers to entities that are known to the addressee but have not been talked about before, that is, they are newly introduced into the current discourse. If a speaker, for example, says "I saw your father yesterday", it is unlikely for the speaker to assume that the addressee had no previous knowledge of his father. The point is that the speaker rather assumed that the entity 'father' had currently not been activated in the addressee's consciousness, that is, he had not been thinking about his father at the moment.

The term 'new' can also be understood as referring to entities that the listener is totally unfamiliar with, that is, the listener has no knowledge of these entities at all. Chafe's notion of givenness and this one are called discourse-old/new and hearer-old/new information, respectively, by Prince (1992). Prince distinguishes a third notion of givenness, namely that of inferable information, and defines it as entities the speaker evokes in the discourse while assuming that the addressee "can infer the (discourse-)existence of certain other entities, based on the speaker's beliefs about the hearer's beliefs and reasoning ability" (1992: 304). Prince leaves open the precise treatment of inferable information, confused about the fact that it is technically hearer-new, but since it is inferable from the preceding discourse and has a discourse-old 'trigger' it is in a way also discourse-old (1992: 307). She eventually suggests that inferable information may be treated as discourse-old.

Prince introduces a matrix capturing and cross-tabulating the hearer- and discourse-status of an entity. As can be seen in Table 3.1, information that is hearer-new and discourse-new is called 'brand-new', hearer-old and discourse-new information is called 'unused' and 'evoked' information refers to entities that are hearer-old and discourse-old. [3] For example, in the sentence *Tom bought a book on apples and read it in the garden*, *Tom* is discourse-new and (presumably) hearer-old (i.e. unused), *a book* is discourse-new and hearer-new (i.e. brand-new) and *it* is discourse-old and hearer-old (i.e. evoked).

3 For more information on Prince's taxonomy of information statuses see Prince (1981).

Table 3.1: The distinction between given and new information according to Prince (1992).

	hearer-old	**hearer-new**
discourse-old	evoked	[non-occurring]
discourse-new	unused	brand-new

Adapted from Prince (1992: 309).

A more recent conceptualization of givenness is provided by Gundel and Fretheim (2004) who distinguish between referential givenness and relational givenness. The concept of relational givenness

> involves a relation between a linguistic expression and a corresponding non-linguistic entity in the speaker/hearer's mind, the discourse (model), or some real or possible world, depending on where the referents or corresponding meanings of these linguistic expressions are assumed to reside. (Gundel/Fretheim 2004: 176)

Examples of this notion of givenness include Prince's (1992) discourse-old/new and hearer-old/new statuses, the activation and identifiability statuses of Chafe (1994) and Lambrecht (1994) and the cognitive statuses of Gundel et al. (1993). Gundel et al. assume that different personal and demonstrative pronouns signal different cognitive statuses (attention states in the addressee's mind) and suggest that these statuses are necessary for explaining the relation between these referring expressions and their use and interpretation in discourse.[4] Gundel and Fretheim (2004) elaborate on this idea and state that these referential statuses are "uniquely determined by the knowledge and attention state of the addressee at a given point in the discourse. The speaker has no choice in the matter" (178).

Relational givenness, on the other hand, involves the partition of the sentence into two parts X and Y, where X is what the sentence is about and Y is what is predicated about X (Gundel/Fretheim 2004: 177). The distinction between these two parts of a sentence has a long tradition and various terms and definitions have been suggested in the literature, including the distinctions between theme-rheme (e.g. Halliday 1967; Firbas 1964), presupposition–focus (e.g. Chomsky 1971; Jackendoff 1972), topic-comment (e.g. Hockett 1958; Hornby 1971) and topic-focus (e.g. Erteschik-Shir 2007).[5] These concepts will be discussed in some more detail in the next section. I will use the terms 'topic', 'focus' and 'comment' to refer to these concepts. As for the definition of givenness, the present study works with the concept of discourse-old/new information, that is, entities will be identified as 'new' if they are newly introduced into the discourse.

4 In their theory of givenness, Gundel et al. (1993) introduce the framework of the so-called Givenness Hierarchy, where referring expressions are organized according to their cognitive statuses. This inclusive hierarchy includes items that are 'in focus', 'activated', 'familiar', 'uniquely identifiable', 'referential' and 'type identifiable'.

5 Other linguists have also commented on the importance of distinguishing between relational and referential givenness. See, for example, Halliday (1967).

3.1.3 Topic - focus

The term 'topic' has first been proposed by Hockett (1958) to capture a linguistic notion that has some similarity to the syntactic notion of subject but which is more difficult to define. He notes that "the most general characterization of predicative constructions is suggested by the terms 'topic' and 'comment' for their ICs [Immediate Constituents]: the speaker announces a topic and then says something about it" (1958: 201). This basic understanding of the notion of topic in terms of aboutness has already briefly been touched upon in the section on word order above. Discussing the examples in (3.1), *Tom loves apples* and *Apples Tom loves*, we have seen that the two sentences differ with respect to their topics: while the first sentence is about Tom, the second sentence is about apples.

Topics are normally associated with given or known information, while foci or comments provide new information about the topics (cf. e.g. Hockett 1958: 201; Hornby 1971: 1976; Halliday 1967: 212; Reinhart 1981).[6] There is a large body of research on the notion of topic, proposing different definitions from very different perspectives. From the perspective of linear order, topics are identified with initial position, constituting an addressee-oriented strategy as the addressee is thus provided with an easily accessible and familiar referent (Chafe 1976; Li/Thompson 1976; Halliday 1967; Seoane 2006). Syntactically, topics are often associated with the subject (e.g. Givón 1976; Li/Thompson 1976; Reinhart 1981; Lambrecht 1994; É. Kiss 2001). Phonetically, topics have often been associated with the non-stressed element in the sentence (e.g. Chomsky 1971; Jackendoff 1972). Defined in functional terms, topics can be described as discourse features since they determine the theme of the discourse and - at least in English - the previous discourse is often needed to safely identify the topic of a sentence (cf. the sentences in (3.1) and the discussion about them). From a cognitive perspective, topics can be described as the centre or focus of the speaker's attention (e.g. Schachter 1973). Furthermore, it has been claimed that there tends to be a correspondence between topicality and definiteness (e.g. Kuno 1972).

Topics can be marked explicitly in a sentence. This includes morphological marking, such as the topic marker *wa* in Japanese or the marker *nŭn* in Korean (Kuno 1972; Primus 1993), and syntactic marking, which includes the constructions that are subject to the present study. They will be discussed in some detail later in this chapter.

Languages do not only differ in the way topics are realized or marked, but they can also be more fundamentally distinguished by the importance of the notion of topic

6 Topics do not necessarily have to refer to old information, that is, referents that are mentioned for the first time in discourse can be topics. Such topics often involve some deictic expression and exist or appear in the current situation. Discussing this problem, Erteschik-Shir (2007: 18) considers the example sentence *That chair is ugly*, where the topic *the chair* has not been part of the preceding discourse. As a solution to this problem she suggests to make a distinction between 'old' and 'given' information, stating that "old means that the referent has been mentioned in the conversation; given, however, means that the hearer has the referent in mind" (2007: 18). Erteschik-Shir then concludes that "topics must be given" (ibid.). Note that her distinction between 'old' and 'given' seems to resemble Prince's (1992) distinction between discourse-old/new and hearer-old/new statuses. Recall, furthermore, that both these statuses are defined as referential givenness by Gundel and Fretheim (2004), while the topic is associated with yet another sense of givenness, relational givenness. This further illustrates the difficulties that are involved in finding a precise definition of the term 'topic'.

as opposed to the notion of subject in the structuring of sentences. That is, there are languages in which basic sentences are structured around subject and predicate, as in (3.1), where *Tom* is the subject and *loves apples* the predicate. English, French and Indonesian are examples of these so-called subject-prominent (SP) languages. In so-called topic-prominent (TP) languages, on the other hand, the basic sentence structure is determined by the topic. Before examining more carefully the major differences between these types of languages, a description of the term 'focus' will be provided.

Like topics, foci have also been defined in various ways from different perspectives in the literature.[7] What all the different definitions and uses of the linguistic term have in common, however, is that they all have to do with the highlighting of constituents, making them stand out from the other constituents in the sentence. Focus is put on a constituent to highlight the information it carries, to introduce new information into the discourse, to re-introduce a referent after a longer gap of absence, to contrast one piece of information with another, or to shift the addressee's attention to another entity or topic. This can be done in various ways, including phonetic, morphological and syntactic means.

Foci have often been described as the new information in relation to the topic (e.g. Vallduví 1994; Miller 2006). Many (but not all) languages use stress to mark foci and if both topic and focus are stressed then the focus receives the most prominent stress (e.g. Halliday 1967; Cinque 1993: 257; Erteschik-Shir 2007: 30; Drubig/Schaffar 2001). Cinque (1993) elaborates on this observation and argues that the focus receives not only the most prominent stress but is also the most deeply embedded constituent in terms of surface sentence structure. Chomsky (1971) and Jackendoff (1972) define the focus as the non-presupposed information in the sentence, with presupposition being defined as "the information in the sentence that is assumed by the speaker to be shared by him and the hearer" (Jackendoff 1972: 230). From a pragmatic perspective, foci have been associated with speakers' intentions and defined as the information that is most important or salient in the current context and that the addressee should add to his knowledge (e.g. Erteschik-Shir 2007: 38ff.).

Gundel and Fretheim (2004) describe foci in similar terms, yet they take the aforementioned features of foci as actually describing two different types of focus. There seems to be growing consensus among researchers of information structure that two types of focus must be distinguished in terms of form and interpretation (cf. e.g Halliday 1967; Rochemont/Culicover 1990; É. Kiss 1998). Information focus (or presentational focus, predicate focus, rheme, wide focus, comment) is interpreted in relational terms referring to information that says something about the topic, that is, it merely marks new information and usually occurs clause-finally. Contrastive focus (or identificational focus, operator focus, narrow focus), on the other hand, is often found clause-initially. It refers to that information in the clause the speaker calls to the addressee's attention.[8] For illustration of the difference between information focus

7 For a comprehensive overview of theoretical approaches to the linguistic term 'focus' see, for example, Winkler (1997) and Miller (2006).

8 In addition to Gundel and Fretheim (2004), see É. Kiss (1998, 2001) for a thorough discussion of the distinction between information focus and contrastive focus. Lambrecht (1994: 221ff.) distinguishes between three types of focus. His predicate focus correlates with Gundel and Fretheim's (2004) infor-

and contrastive focus consider the following two examples of fronting constructions. Capital letters indicate prosodically prominent items.

(3.2) We have to get rid of some of these clothes. That COAT you're wearing I think we can give to the Salvation ARMY.

(3.3) A: Which of these clothes do you think we should give to the Salvation Army?

B: That COAT you're wearing (I think we can give away).

(Gundel/Fretheim 2004: 182-183)

The two examples have in common that they both have a prosodically prominent preposed object (*that coat*), yet they differ in the information status of the preposed element. In (3.2), the coat is the topic, possibly contrasted with some other entities of the set of clothes (contrastive focus). In (3.3), on the other hand, the coat is part of the information focus, providing new information about the topic 'clothes that should be given to the Salvation Army'. These examples also show that there are basically two different types of fronting construction, which are often confused in the literature: object fronting, as in (3.2), and focus fronting, as in (3.3).[9]

In addition to prosodic and syntactic marking, foci can also be marked morphologically. In English, for example, foci can be marked with the help of the focus particles *even*, *only*, *also* and *just*. Differences in their distribution and placement across varieties of English have been identified in the literature (cf. e.g. Parviainen 2012; Lange 2007, 2012; Fuchs 2012). Functionally, all these particles have in common that they highlight a certain constituent in the sentence, making it salient. But they may be used for very different reasons, for example, to highlight new information, mark a transition or to help signal the speaker's attitude.

3.1.4 Topic-prominent languages

In topic-prominent (TP) languages, such as Chinese, Japanese, Korean and Tagalog, basic sentences are structured around topics rather than subjects. In these languages, the topic-comment[10] structure can be seen as part of the repertoire of basic sentence

mation focus. Contrastive focus or his argument-focus structure applies to sentences "in which the focus identifies the missing argument in a presupposed open proposition" (1994: 222). The third type of focus, which Lambrecht defines as sentence focus, involves sentences where "the focus extends over both the subject and the predicate" (ibid.). An example would be the answer to the question 'What happened?'.

9 Some authors represent the referents of discourse metaphorically as a set of file cards (Reinhart 1981; Lambrecht 1994; Erteschik-Shir 2007). The cards are organized in such a way so that the most activated constituents are on top of the stack (or in the front section if we compare them with a drawer with file cards). These constituents represent the potential topics for the ensuing discourse. Foci can also be placed on top of the stack. For more information on this interaction between topic, focus and syntax see Erteschik-Shir (2007: 43-45).

10 The term 'comment' is here and throughout the present study used to refer to that part of a sentence that says something about the topic (i.e. information focus).

types, that is, the topic function is integrated into the basic syntax of the sentence (Li/Thompson 1976: 471). In subject-prominent (SP) languages, on the other hand, basic sentences have a subject-predicate structure. This does not mean that SP languages are devoid of topic-comment structures; it rather means that the syntactic category 'subject' is structurally more important than the categories of information structure (i.e. topic and comment). In SP languages, the topic is either expressed by the subject, which then constitutes an unmarked topic as in sentence (3.1a) above, *Tom loves apples*, or it is highly marked and set off from the rest of the sentence, as is the case, for example, in the sentences in (3.4), where constructions with *as for* and *as far as x is concerned* are used to mark the topic (further possibilities include *regarding x*, *concerning x*, among others).[11]

(3.4) a. As for apples (topic), my grandma has a huge tree in her garden.

b. As far as that tree is concerned, the apples are delicious.

In addition to these structures, topics can be syntactically marked in English by means of left dislocation constructions (e.g. *Tom, he loves apples*).

The basic ideas of the topic-comment structure can well be illustrated by the so-called 'double-subject' construction, illustrated in (3.5) with an example from Mandarin.[12] Japanese provides another illustrative example. In this language topics are typically marked with the morpheme *wa* and subjects with the morpheme *ga*, as in (3.6).

(3.5) *Nèike shù yèzi dà.*
that tree leaves big
'That tree (topic), the leaves are big.'

(3.6) *Sakana wa tai ga oisii.*
fish TOP red snapper SUBJ delicious
'Fish (topic), red snapper is delicious.'

(Li/Thompson 1976: 468)

In these examples we have both topic and subject, yet the positions are filled by different elements in the sentence. The topic is placed in initial position, which is typically the case, and it has no selectional relationship with the verb but the subject does. The double-subject construction does not occur in SP languages, but they have been found in all TP languages investigated by Li and Thompson (1976: 468), and are indeed quite common, for example, in Mandarin. In Japanese, both the subject and the topic are marked by particles (*ga* and *wa*, respectively).

11 Note that the correspondence between subject and unmarked topic (in declarative sentences), which is so well-established in present-day English, has only developed after the loss of V2. Old English (a V2 language) imposes no restrictions upon the syntactic category that can be placed in initial position of the clause and can encode unmarked topics. After the loss of V2, however, the initial position is reserved for nominal subjects and, thus, it is this syntactic category that came to be correlated with unmarked topics in present-day English (in thematically unmarked clauses, that is) (cf. Seoane 2006).

12 Also see the discussion of information packaging in Mandarin in section 3.4.

Chafe (1976) even goes so far as to suggest that the topic in TP languages refers to something that cannot be translated into any plausible equivalent in SP languages. He argues that the standard English translations with *as for* [topic] for a Chinese topic is not correct because "Chinese seems to express the information in these cases in a way that does not coincide with anything available in English. In other words there is no packaging device in English that corresponds to the Chinese topic device, and hence no fully adequate translation" (1976: 50). Addressing the question of what Chinese topics are, Chafe argues that they set "a spatial, temporal, or individual framework within which the main predication holds" (ibid.). This idea is well illustrated by the double-subject construction in (3.5). The topic *the tree* sets the scene or, more precisely speaking, defines the location in which the predication holds. This is partially also true for topics in SP languages, where topics are defined as those expressions whose referents the sentence is about. This can also be seen as some kind of scene setting. What is different with the Chinese topic, however, is that it is much more loosely linked to the rest of the sentence.[13]

The typological distinction between TP and SP languages is relevant for the present study in so far as some of the L2 English speakers investigated have a topic-prominent L1. This, in turn, may have structural implications for the English variety they speak. Cantonese and Mandarin, two topic-prominent Chinese varieties, are spoken by the majority of the populations in Hong Kong and Singapore, respectively. Furthermore, many Philippine languages are topic-prominent, which might impact on the structure of Philippine English.

This typological difference raises interesting questions. Do different information-structuring principles in a learner's L1 affect the structure of the English variety that emerges? How does an L2 English speaker's prior knowledge of his L1 shape the acquisition of fronting constructions, left and right dislocation constructions, existential clauses and clefts in English? Can we find differences across English varieties in the organization of information in a sentence which can be attributed to the typological make-up of the background language(s)? That is, can systematic differences be identified between L2 English speakers with a topic-prominent L1 and L2 English speakers who have no topic-prominent L1?

3.2 Information-packaging constructions

Present-day English has only a small number of syntactic structures that deviate from the canonical word order, organizing the information in the sentence in a marked way (cf. Speyer 2012). Such information-packaging constructions often have basic (or canonical) counterparts from which they differ syntactically but not in terms of truth conditions. Recall the basic sentence in (3.1a) above, *Tom loves apples*, and the

13 Sentences such as the ones in (3.5) and (3.6) are sometimes also referred to as 'hanging-topic' constructions. For more on these structures see sections 3.2.3 and 4.3.

non-canonical structure with the preposed object *apples* in (b), repeated in (3.7a). The two sentences do not differ in what they say but in the way they say it, that is, in the way the informational content is presented. The same proposition can also be expressed by the following non-canonical syntactic structures:

(3.7) a. Apples Tom loves. [fronting]
b. It's Tom who loves apples. [IT-cleft]
c. What Tom loves is apples. [pseudo-cleft]
d. There is Tom who loves apples. [existential]
e. Tom, he loves apples. [left dislocation]
f. He loves apples, Tom. [right dislocation]

The felicitous use of these constructions is pragmatically constrained, dependent, for example, on the communicative context, the speaker's intentions and the addressee's knowledge. The structures differ in the way the information is organized, that is, the distribution of given and new information and the representation of topic and focus.

In English, fronting constructions require that the preposed element be discourse-old, acting as a link to the preceding discourse, when we are dealing with a preposed topic (e.g. Reinhart 1981; Birner 1997; Ward/Birner 2004). A preposed (information) focus, on the other hand, has no such restriction (Gundel/Fretheim 2004).

Constituents and the information they carry can also be highlighted by means of cleft constructions. In IT-clefts, the focal accent normally falls on the clefted constituent (the post-copular element *Tom* in (3.7b)) and often involves contrast (*It's Tom who loves apples and not Tina.*). Typically, the clefted constituent carries new information, while the cleft clause expresses presupposed information (e.g. Prince 1978; Ward et al. 2002). In pseudo-clefts, the cleft clause does also normally express presupposed or old information (in the sense that it can be assumed to be in the addressee's consciousness at the time of speaking) while the highlighted element expresses new information. Yet, there are also types of cleft construction where both constituents express new information (cf. informative-presupposed IT-clefts discussed below) or where the information in the cleft clause is (represented as) new and the information in the clefted constituent is old.

In existential *there*-constructions, a constituent is highlighted by the use of a special construction and by being in (near-)final position (Miller 2006). The construction is typically used to introduce hearer-new entities into the discourse, that is, the postverbal element represents information that is hearer-new and focal. Another non-canonical syntactic structure that is often used to introduce a new entity into the discourse is left dislocation. The dislocated element is not only highlighted but also marked as topic and it can represent discourse-new and discourse-old information (Ward et al. 2002).

Finally, the information structure of right dislocation is quite straightforward because its form constrains its information-packaging function: the co-referential pronoun represents familiar information and from this follows the discourse-old information status of the right-dislocated element. Note that it is not only required that the right-dislocated element is discourse-old, but it must also represent topical information (Ward et al. 2002).

After this brief overview of the information structure of fronting, cleft constructions, existential clauses and dislocation, the following subsections will describe in some more detail the nature of these information-packaging constructions and the variants that have been included in the present analysis. It should be noted that there are, of course, more or less typical representatives of each category. Some structures are clearly used for information-packaging purposes while others might predominantly serve other functions. Expanded right dislocation constructions are such a case in point, which mainly serve a reinforcing function. It was decided to include these structures in the study nevertheless because they can be regarded as peripheral representatives of right dislocation (also see Durham 2011) and the investigation of these structures yield interesting results. Furthermore, it is of course often the case that structures may serve different functions in discourse at the same time, with some functions being more and others less prevalent.

The structure of the following subsections is as follows: first, the two types of dislocation will be discussed, followed by fronting constructions, existential *there*-constructions and finally various types of cleft construction.

3.2.1 Left dislocation

One of the first authors commenting on dislocation constructions was probably Otto Jespersen (1933), calling it 'extraposition': "[a] word or group of words is often placed by itself, outside the sentence proper, in which it is represented by a pronoun" (1933: 95). That is, prototypical dislocation constructions consist of a noun phrase in a peripheral position which is co-referential to a pronoun in the core of the clause. Two main types of dislocation can be distinguished, namely left and right dislocation, with the noun phrase being placed at the left or right periphery of the clause proper, respectively.

Left and right dislocation constructions are almost exclusively conversational phenomena and occur only rarely in formal registers (cf. e.g. Halliday 1967: 241; Givón 1979: 229; Givón 1983: 347; Geluykens 1992: 34; Lambrecht 1994: 182). They "are well suited to the needs of conversation" (Biber et al. 1999: 957) as they facilitate the online production and processing of an utterance by breaking it down into smaller chunks. According to Carter and McCarthy (1995), the dislocated elements are optional, but if they are used they "carry important interpersonal functions" (1995: 151). In what follows, these functions will be outlined for left dislocation constructions. In addition, the construction's form and realizational variants included in the present study will be discussed. Right dislocation constructions will be the subject of the second subsection.

A typical example of left dislocation (LD) is given in the following extract, in which two girls talk about teaching. The dislocated noun phrase is marked in bold, the co-referential pronoun with [].[14]

14 For clarity reasons, most of the mark-up of the original transcriptions has not been included in this and the following examples. Here, only short <,> and long <„> pauses are indicated.

(3.8) A: You don't get good teachers over here

B: There is no professional approach towards teaching <,,>

A: But that really we learnt in Bombay <,>

How to be a professional in your jobs <,,>

B: That's it

A: **Shireen** <,,> [she] was excellent at that <,,>

She was very good <,,>

She would just you know even even when there was bickering between teachers and all that <,>

If ever we went to her with our problems <,> *ke* so and so is you know she used to say just rise above all that <,,>

(ICE-IND:S1A-003)

The referent *Shireen* has not been mentioned before and is here newly introduced into the discourse by speaker A, who establishes her as the topic by means of a LD construction. The following utterances are all about her, where she is pronominally referred to. This example sentence thus illustrates one of the major communicative functions of LD constructions, namely topic-promotion or referent-introduction (cf. e.g. Geluykens 1992: 51, Gregory/Michaelis 2001: 1680). That is, the speaker establishes the referent first, making sure that the interlocutor can follow (cf. the pause <,,> after the introduction of the topic *Shireen*), and then continues with some information or comment on the referent. LD tokens are thus well suited to the interactional nature of conversation, exhibiting Lambrecht's (1994) *Principle of Separation of Reference and Role*: "Do not introduce a referent and talk about it in the same clause" (1994: 185).[15]

Further discourse functions have been identified in the literature, including the so-called 'simplifying' function and the 'poset' function introduced by Prince (1998).[16] 'Simplifying' for Prince (1998: 286) refers to the discourse processing of DISCOURSE-NEW entities. Discussing LDs whose preclausal element evokes a discourse-new referent, she looks at the position the initial element would occupy if the sentence was in canonical form and finds that in all of her examples "the NP would canonically be in a position that is strongly disfavored for NPs evoking Discourse-new entities" (ibid.), namely subjects, possessives and embedded items (the object position would be the favoured position). 'Simplifying' LDs thus serve to simplify the processing of such items by removing the discourse-new element from a disfavoured position in the clause and "creating a separate processing unit for them" (ibid.). In addition to Prince's interpretation, the term 'simplifying' can also be understood in a more syntactic way. LD constructions serve both speaker and hearer as they facilitate the

15 For empirical evidence of Lambrecht's principle with regard to the LD construction see Kuzar/Netz (2010).

16 Prince (1998) actually distinguishes three types of LD constructions. Along with the 'simplifying' and 'poset' LD she discusses so-called 'resumptive pronoun topicalization' LDs which, for her, are fronting constructions "in disguise" (1998: 291). This third type is similar to resumptive pronoun relative clauses, which will not be discussed in the present study. Hence, only the two former types of LD will here be considered.

planning and processing of syntactically complex sentences by breaking them down into smaller units (cf. e.g. Biber et al. 1999: 138; Ward et al. 2002: 1409). Complex preclausal elements may, for example, consist of a noun phrase modified by a relative clause, as illustrated in (3.9).[17]

(3.9) a. **The people we were staying with** [they] <,,> uh cooked us a traditional Normandy dinner <,,> (ICE-GB:S1A-009)

b. **The only other Mason who was a garage guy in Crossgar** Dad absolutely hated [him] (ICE-IRE:S1A-005)

c. But **this lady that was with her** <,> [her] husband took a night off (ICE-IRE:S1A-055)

Prince's (1998) second type of LD construction - 'partially ordered set' LDs or 'poset' LDs, for short - involves a set relation between the initial noun phrase and some entity in the preceding discourse, for example, *is-a-part-of* or *is-a-subtype-of* relations. An example of this type of LD is given in (3.10).

(3.10) B: And how many brothers and sisters you have?

A: Uh I have that uh <,,> two brothers <,>

So including me we are three <,>

And uh two sisters <,>

So <,> I'm the youngest I am the youngest one

B: Uhm.

A: And uh my eldest brother is <,> uh taking care of this uh uh uh <,>

B: Agriculture <,>

A: Yeah <,>

And basically we are agriculturists <,>

And **my middle brother** [he] was a school master <,>

And very recently he retired as a head master of a <,> primary school

B: Uhm <,>

A: And **my two sisters** [they] are married to some uh <,> village people and they are alright after all <,> (ICE-IND:S1A-076)

Speaker A is asked about her brothers and sisters. The initial elements of the two LD tokens, *my middle brother* and *my two sisters*, are obviously members of this set of brothers and sisters.

As far as the annotation of the data is concerned, the relevant dislocation constructions have been coded manually in the present study. Along with the prototypical

17 Complex LD tokens, their frequency and distribution across varieties of English will be discussed in some more detail in the analysis section 4.1.

constructions with a detached noun phrase and a co-referential personal pronoun (cf. the sentence in (3.8)), further less typical sentences have been marked as LD tokens. These include questions, as in (3.11a), and constructions with co-referential demonstrative pronouns, such as *this* and *that*, in the core of the clause, as in (3.11b).

(3.11) a. **One box** is [it] enough? (ICE-SIN:S1A-006)

b. **Lashkar** [that]'s the capital of Tibet (ICE-NZ:S1A-024)

In some LD sentences we find variation with the fronted elements. In addition to the prototypical noun phrase we find gerunds, finite and infinite clauses, as in (3.12a), (3.12b) and (3.12c), respectively. It is certainly debatable whether these sentences should be treated as dislocation constructions at all. Yet, it was decided that they sufficiently resemble the LD structure proper to be included in the analysis: the pronoun in the core of the clause is co-referential with the fronted element and can equally well be left out.

(3.12) a. Yes **clapping and singing** he's against [it] three hours non-stop (ICE-GB:S1A-068)

b. **How the <,> standard of <,> uh women has been exposed there** <,> I don't like [this] (ICE-IND:S1A-088)

c. Because **to make a pizza** [it] costs you like ten p and you sell it for fucking two quid (ICE-IRE:S1A-010)

Another variant of dislocation construction which is part of the analysis involves a three-step realization, that is, after establishing the referent, the speaker is interrupted by the hearer or rather awaits their consent to continue. An example of this type of LD token is given in (3.13).

(3.13) A: Uhm **Fergie and Diana**

B: Yeah.

A: [They]'re on their way to a Palace uhm reception

(ICE-GB:S1A-041)

Such sentences underline the fact that LD constructions are to be situated at the interface between syntax and discourse. That is, they are not to be seen as purely syntactic phenomena, but the context in which they occur plays an important role as well. This also becomes obvious in their main function, namely topic promotion, as mentioned above (also see the detailed analysis of LD tokens in section 4.1).

Furthermore, a number of rather non-standard LD tokens have been included in the analysis. They are part of the study because, for one, they reflect the nature of the data - spontaneous spoken interaction - and, secondly, they contain interesting features that are characteristic of the speech of L2 English speakers. Consider, for example, the sentences in (3.14), taken from the Jamaican English data. What makes these examples exceptional is the use of the pronouns *them* and *him* instead of the standard expressions *they* and *he* as well as the omission of the copular *be* in sentence (a).

(3.14) a. Yeah oh **her lips** [them] sweet big but sweet (ICE-JAM:S1A-036)

b. **The other bredrin that work with me** [him] gone out too for a week (ICE-JAM:S1A-037)

The use of the same pronoun for subject, object and possessive function (i.e. *me*, *him* and *them*) in Jamaican English is due to influence from Jamaican Creole, which does not mark case or gender on pronouns. Yet note that such items occur rather rarely in Jamaican English and pronoun use is largely based on Standard English forms (Deuber 2014: 108). The same is true of zero copular. In Jamaican Creole, it is the dominant form before adjectives, in progressive forms and the *going to*-future; in Jamaican English, zero copular also occurs, especially with verbal predicates, but it is much rarer and *be* clearly predominates in all environments (Deuber 2014: 86-87).

Finally, the sentences in (3.15) exemplify a type of construction that has NOT been included in the present investigation.

(3.15) a. **My Dad** *is* uhm he's very he's not [he] wouldn't be extremely well-educated like he did an apprenticeship (ICE-IRE:S1A-011)

b. **Germany** *is is* uh [it]'s better than what I expected you know uhm (ICE-JAM:S1A-072)

These sentences reflect two major factors that constrain conversation, namely online processing and interactiveness. The speaker in (3.15a) starts out with the first three words of a canonical sentence, *My Dad is*. He then becomes hesitant and after a filled pause (cf. hesitation marker *uhm*) and some false starts (cf. *he's very he's not*) eventually finishes the sentence by repeating the subject of the clause (*My Dad*) in the form of the co-referential pronoun *he*. This makes the construction very similar to the LD construction proper. The same is the case with the example sentence in (3.15b). Here, the speaker is interrupted by overlapping speech while uttering the word *Germany*. He then continues after a filled pause with the co-referential pronoun *it*. After some consideration it has been decided to exclude such examples from the present study because the speakers start out with a canonical sentence and the resumptive pronoun follows only after 'false starts' and pauses, which makes an interpretation of the constructions' functions and their classification difficult.

3.2.2 Right dislocation

This section describes the nature and functions of right dislocation constructions, or RDs for short. In this construction, a constituent is taken out of the core of the clause and placed clause-finally. In the position where the constituent would normally occur we find a co-referential pronoun. Typical examples are given in (3.16). The right dislocated elements are marked in bold print, the co-referential pronouns with [].

(3.16) a. [He]'s brilliant **your dad** (ICE-GB:S1A-042)

b. So anyhow what are [you] doing tonight <,> **yourself and Nora** <,> Ciara? (ICE-IRE:S1A-089)

It is rather difficult to precisely specify the discourse functions of RD constructions and researchers in the field disagree in this respect. The two major functions that have repeatedly been suggested in the literature are illustrated in the examples in (3.16). RD tokens often contain an affective or emotive dimension, serving to establish a bond with the interlocutor, or they are associated with some form of evaluation (cf. e.g. Visser 1963: 54; Aijmer 1989; Carter/McCarthy 1995; McCarthy/Carter 1997; Timmis 2010). This function is illustrated by the example in (3.16a) where the referent *dad* is considered to be brilliant, undoubtedly an evaluation.

Furthermore, RD tokens can have a clarifying or disambiguating function. Starting out with a pronoun in the core of the clause, the speaker realizes that the referent needs further clarification (cf. Aijmer 1989; Biber et al. 1999: 957; Timmis 2010). Example (3.16b) is very likely such a case. The pronoun *you* does not sufficiently clearly indicate the persons meant. Hence, after a pause the speaker adds the noun phrase *yourself and Nora* for clarification.

Given these discourse functions, RD constructions are valuable devices for conversations as they serve well the reciprocal and dialogic nature of ongoing interaction. For one, they allow the speaker "to cope with planning pressure, and at the same time to convey some fairly complex messages" (Biber et al. 1999: 1072). Furthermore, they go well with another feature of conversation, namely the expression of feelings and attitudes. A third discourse function - that of emphasizing the proposition of the clause - will be discussed in more detail below because it is a minor function applying to specific types of RD tokens only.

The information status of the detached element in RD constructions is very different from that in LD constructions. While the latter type of structure may be used with both referents referring to old information and referents whose topic status is not yet established in the discourse, in RD constructions the detached constituent is always already highly salient since it has been pronominally referred to before in the sentence.

In addition to the prototypical examples given in (3.16), a number of further variants of the RD construction have been included in the present study. Dislocated elements are typically placed at the end of the utterance, but they may also occur in the middle of the utterance, sometimes even right after the co-referential pronoun. Examples are given in (3.17).

(3.17) a. And people [it] was as strong as it was like **the tide** that you could get knocked off your feet <,> if you were like uh unsteady or something (ICE-PHI:S1A-007)

b. And there were still hundreds of people on it but [it] was so big **this boat** that you didn't didn't meet them (ICE-GB:S1A-021)

c. I'm reading [it] **Treasure Island** at the moment to my son (ICE-GB:S1A-013)

As with left dislocation, we also find questions and constructions with co-referential demonstrative pronouns, as illustrated in (3.18).

(3.18) a. How was [it] **your initial reaction towards each other**? (ICE-GB:S1A-001)

b. [That]'s a nice area isn't it **Leatherhead**? (ICE-GB:S1A-081)

The dislocated element is most commonly realized in the form of a noun phrase, but there are also instances of final demonstrative pronouns and clauses (finite and non-finite), as can be seen in the sentences in (3.19).

(3.19) a. [It]'s a cultural thing **that** you know (ICE-GB:S1A-043)

b. When was [it] made **this**? (ICE-GB:S1A-019)

c. I mean it sounds like there's uh technical <,> things like release work and [it] sounds like some kind of therapy **that you're doing** (ICE-GB:S1A-004)

d. My God [it] makes me cry though **watching things like this** (ICE-GB:S1A-042)

e. [It]'s a hundred and fifty pound job **to replace a door** (ICE-GB:S1A-007)

RD structures with co-referential pronoun *it*, as in (3.19d) and (3.19e), need to be distinguished from superficially very similar extraposition constructions. Examples of the latter type of construction are given in (3.20). Sentences involving extraposition often have an infinitive clause or a content clause as the right-dislocated element, as exemplified in (a) and (b), respectively. Participial clauses/gerunds, on the other hand, are "at best very marginal" in extraposition constructions (Ward et al. 2002: 1407). Furthermore, sentences like the one in (c) are quite common. They have an exclamative function, usually asserting that a certain situation is remarkable (Michaelis/ Lambrecht 1996: 228).

(3.20) a. I know they don't but it seems pointless to have it without them (ICE-GB:S1A-068)

b. And then it turns out that it is not the PNP that is doing this but uhm Mr Earlston Spencer <,> uhm (ICE-JAM:S1A-068)

c. It's just amazing the way she's so quick at picking up the music <,> (ICE-GB:S1A-091)

There are other types of extraposition, but in my opinion RD constructions are most difficult to demarcate from these three structures. However, since there are differences in prosody, semantics, syntax and information structure between RD and extraposition, various tests may help to disambiguate unclear cases (cf. Ward et al. 2002: 1413-1414). First, in contrast to extraposition, the dislocated element in RD tokens is almost always a distinct intonational phrase, that is, we often find a pause before the dislocated element (filled or unfilled). Second, the pronoun *it* in extraposition sentences is non-referential (dummy *it*) while in RD the pronoun is referential. In the latter case, *it* often refers back to something that has been mentioned in the preceding discourse. Consider, for example, sentence (3.19e), repeated in (3.21) with some more context.

(3.21) C: Why don't you uhm replace one of the back doors here and use the pane from that

A: Well if I ever have to replace a <,> back door I shall do so

C: Well they're rotten now Dad so one day you'll have to

A: Well I'm just putting that off for as long as possible

B: They'll fall out into the garden one day

A: **It's a hundred and fifty pound job to replace a door**

(ICE-GB:S1A-007)

Looking at the sentence in isolation it would probably rather be treated as an example of extraposition because dislocated infinitive clauses occur much more frequently in this type of construction than in RD tokens. Yet, if the preceding discourse is taken into account I would argue that the pronoun *it* has referential meaning, referring back to what has been said before about replacing the back door. This example is therefore classified as an instance of RD in the present study.

Furthermore, there is a difference between the information status of the right-dislocated and extraposed constituents. While right-dislocated elements always refer to discourse-old information, extraposed elements may contain discourse-new information. These tests have been applied to ambiguous sentences in the present study. If an unclear case could not be disambiguated with any of these tests, it was excluded from the analysis.

Finally, another specific type of RD construction included in the present analysis contains an operator in the dislocated element. Examples such as (3.22a) and (3.22b) have been reported in the literature under various names, for example, 'expanded right dislocation' (Durham 2011)[18], 'declarative tag' (Biber et al. 1999: 139) or 're-inforcement tag' (Carter/McCarthy 1995: 150). Example (3.22c) constitutes an even more specific type of expanded RD token. Here, the dislocated noun phrase and the operator are preceded by the particle *so*. This type of RD structure, which I labelled '*so*-tag', seems to be a characteristic feature of Irish English because quite a number of such tags can be attested for this variety of English but hardly any for the other varieties.

(3.22) a. [It] was just you know the market in Cambridge **it was** yeah
(ICE-GB:S1A-015)

b. [He]'s going to Urban Dance Squad **Phil Alexander is** (ICE-GB:S1A-100)

c. [That]'s pure selfish **so it is** (ICE-IRE:S1A-050)

d. [I] used to have great craic with them in England <,> **so I used to**
(ICE-IRE:S1A-087)

The expanded type of RD is a historically more recent structure. While the origins of the 'canonical' RD construction can be traced to Old English, the first attestations of the expanded form date to the second half of the nineteenth century (Visser 1963: 54-55). It seems that they are today mainly found in northern dialects of British English (Durham 2011).

18 More detailed information on Durham's discussion of this type of right dislocation is given in the literature review in section 3.3.

The examples in (3.22) illustrate a third function RD constructions may serve in addition to the two major discourse functions mentioned above (disambiguating function and emotive function). Expanded RDs are mainly used to emphasize the proposition of the clause.[19]

The right-dislocated elements, or tags as I will sometimes call them, in expanded RDs may be preceded by a pause, as in example (3.22d), but usually they are directly adjacent to the rest of the clause. In the data there are also a number of examples where the tag is labelled as a separate speech unit. Debatable cases indeed. After some consideration it was decided to nevertheless consider them as instances of (expanded) RD tokens because it is the transcribers of the recorded conversations that make these classifications and some other transcriber may have put a pause there instead. Consider the following examples for comparison. In (3.23), speakers B and C talk about a film C has not yet seen but obviously also does not intend to watch. The 'potential' expanded RD is marked in bold print. Note that we have two speech units. In (3.24), on the other, we have a very similar construction (again marked in bold print) but, here, we do not have two speech units. There is only a pause before the tag.

(3.23) C: I have never seen it

still I really [unclear words]

really it's not my sort of film <,,>

B: But it it's very similar to Nineteen Eighty-Four in that you know there's like the baddies and the goodies <,>

uhm <,> **it's weird**

It really is <,>

Because he gets he's [unclear word] trying to <,> kick the system <,,>

(ICE-GB:S1A-049)

(3.24) B: We had salad and then

A: We only have the small cake these very very small cakes

It tastes good <,> **it really does** uh I mean I didn't eat as much as I should [speaker B chuckles] because I was on a diet

I really was on a diet at that time

(ICE-PHI:S1A-080)

In my opinion these two structures are very similar. Judging from the text and information flow, there could just as well be a pause mark-up after *it's weird* in (3.23). Consequently, I treated both items, that in (3.23) and (3.24), as instances of expanded RD.

19 It is debatable whether these structures should be mentioned under the heading 'information-packaging constructions' at all. They mainly serve a reinforcing rather than information-structuring function. Yet, they are here regarded as a type of RD - although a marginal and less typical variant thereof - and are thus included in the analysis.

3.2.3 Fronting constructions

Fronting, also referred to as topicalization or preposing, is another type of construction in which the canonical word order of English (SVO) is transformed for pragmatic purposes. It refers to "the initial placement of core elements which are normally found in post-verbal position" (Biber et al. 1999: 900). Various lexically governed and non-lexically governed elements can be placed in initial position. Typically, we find a preposed object either in the form of a noun phrase (proper noun or pronoun) or a prepositional phrase. Less typical are clausal fronted objects. The sentences in (3.25) illustrate these realizational options.

(3.25) a. And uhm so I've I've written a few short texts in it like **the Lord's Prayer** I've translated (ICE-GB:S1A-015)

b. But **that** really we learnt in Bombay <,> (ICE-IND:S1A-003)

c. **In Hindi** also we can write <,,> (ICE-IND:S1A-071)

d. But exactly <,> **exactly how it was slotted in** <,> I can't remember quite (ICE-GB:S1A-012)

In addition to fronted objects, we also find fronted complements which are typically realized in the form of noun phrases or adjective phrases, very rarely also as (finite or infinite) clauses. Examples are given in (3.26).

(3.26) a. They had a sports uh <,,> suppliment <,> yeah <,> **extra sport** it was called (ICE-IND:S1A-099)

b. Oh **great** that'd be (ICE-GB:S1A-042)

c. They are very I mean uh <,,> **to make friends** they take a lot of time I don't know why (ICE-IND:S1A-046)

Furthermore, constructions with fronted obligatory adverbials have been included in the present analysis, that is, adverbials that are an argument of the verb. The sentence in (3.27) illustrates this type of fronting construction.

(3.27) **The big mansion** he was in (ICE-IRE:S1A-067)

Optional fronted adverbials, on the other hand, are not part of this study since they are mobile per definition (cf. e.g. Biber et al. 1999: 763). Further constructions that have not been included in the present study involve comment clauses in final position, such as the sentences in (3.28). The reason for their exclusion is the fact that, here, the clause-final expressions *I think* and *I suppose* rather function as some kind of hedges and are "often best analysed as epistemic adverbs expressing speaker attitude, as markers used for discoursal, interactional and interpersonal purposes, or as markers of hesitant phases and mental planning or word-searching phases" (Dehé/ Wichmann 2010: 37).

(3.28) a. She's a student at Saint Martin's **I think** (ICE-GB:S1A-020)

b. Uhm even though we'll be there **I suppose** (ICE-GB:S1A-069)

Finally, the present study will not discuss so-called 'hanging-topic' or 'unlinked topic' constructions, structures in which the preclausal element is only loosely linked to the rest of the clause.[20] The sentences in (3.29) are taken from a dialogue about babies who, speakers A and B agree, outgrow their clothes so quickly that parents usually pass along the clothes to other people. B's sister, however, loves to buy new clothes for B's children so that her cupboard is packed. The second excerpt, also taken from ICE-Singapore, is about C's application for a job at IBM.

(3.29) B: Now my whole cupboard is packed
I think twenty to thirty suits each
A: Wow
B: **Then shoes** they have so many pairs she like must match match them
(ICE-SIN:S1A-048)

(3.30) A: Did you apply for the PR job at uh Marina Mandarin?
C: No I only applied for
In fact **the IBM one** I applied about a year ago
So when they call me up uh yesterday evening about five
I said **IBM** I don't remember applying
(ICE-SIN:S1A-004)

The following example is taken from a conversation about speaker C's plans on making a video with her students, although she is not very good at it herself.

(3.31) C: I cannot handle the camera <,> so video camera so confidently <,> but I am comparatively good at outdoors <,,>
Indoors and closeups I had to take help of my husband and Ganesh <,,>
(ICE-IND:S1A-046)

These sentences clearly show that the fronted elements introduce new topics into the discourse. In more formal English, these topics would probably be introduced by means of an *as for*-construction, for example, *As for shoes then, they have so many pairs*. It has been claimed in the literature that the 'hanging-topic' construction is typical of topic-prominent languages (cf. Lambrecht 2001a; Setter et al. 2010). Consequently, one could expect to find more such items in those English varieties which have a topic-prominent background language. Investigating this hypothesis is certainly worth the effort but has to remain for future research because it is beyond the scope of the present study and a cursory look at my data suggests that larger databases are needed than the 200,000-word corpora I use.

As for the functions of fronting constructions, they are mainly used to establish a link to the preceding discourse and thus create coherence (cf. e.g. Reinhart 1981;

20 Also see the discussion of fronted prepositional phrases in section 4.3 and the difficulty of keeping this type of fronting construction and the 'hanging-topic' construction apart.

Prince 1985, 1998; Biber et al. 1999; Gregory/Michaelis 2001; Ward/Birner 2004; Netz/Kusar 2007). This is probably best illustrated with the example in (3.25b) where the fronted element occurs in the form of discourse-deictic *that*, repeated in (3.32) with some more context to get a better idea of the linking function. Note that through the preposing of *that* the term is also emphasized.

(3.32) A: You don't get good teachers over here
B: There is no professional approach towards teaching <,,>
A: But **that** really we learnt in Bombay <,>
How to be a professional in your jobs <,,>

(ICE-IND:S1A-003)

In addition to items that serve a linking and/or emphasizing function, we also find fronting constructions that are used for contrastive purposes. An example is given in (3.33), an extract from a conversation about favourite books. Speaker E talks about book-hunting in secondhand bookshops in order to find replacements for those of his favourite books that have become dog-eared. Speaker A wants to know why he is doing that.

(3.33) A: Is that because you can't get new copies or you prefer to have?
E: Oh well a lot of these books have been out of print for <,> well not
A: Yes
E: Yeah
A: Hardback
E: No they come back in paperback but I like them in hardback
The favourites
I mean OK I go in Waterstones or wherever and buy paperbacks and read them and probably cast them aside or give them lend them to someone else and forget that you who you've lent them to <,>
But uh **hardbacks** I wouldn't lend to anyone

(ICE-GB:S1A-013)

Note that the fronted element *hardbacks* is not only contrasted with *paperbacks*, but we also find a type/subtype relation with the entity *books*, the general topic of the preceding discourse. Both hardbacks and paperbacks are members of the set 'books'. In section 3.2.1, such partially ordered set or poset relations have already been identified as a discourse function of left dislocation constructions. As far as fronting constructions are concerned, Birner and Ward (1998: 31) claim that ALL tokens are linked to the preceding discourse:

> [T]he various preposing constructions of English form a natural class on pragmatic grounds, in that they are subject to the following discourse constraints: Felicitous preposing requires that the referent or denotation of the preposed constituent be anaphorically linked to the preceding discourse (see Reinhart 1981, Vallduví 1992).

The five poset relations that occur most frequently in Birner and Ward's (1998: 45) data include set/subset, part/whole, type/subtype, greater-than/less-than, and identity.

Given these different functions, two types of fronting construction can be distinguished. In addition to their function, they also differ in terms of intonation and information structure. The fronted element is interpreted as the topic if it receives no focal stress (the element is certainly stressed when fronted, but the degree of stress is less than on the final words of the clause). If the fronted element contains the focus, however, and bears nuclear stress, it does not serve as the topic. Additionally, this latter type (focus preposing) often involves contrast.

3.2.4 Existential *there*-constructions

Existential *there*-sentences typically have the structure

there + *be* + noun phrase.

In these sentences unstressed, non-deictic and non-locative *there* functions as the syntactic subject. It is commonly assumed that it has developed out of the locative adverb *there* (e.g. Bolinger 1977; Breivik/Swan 2000; Ward et al. 2002: 1391; Breivik/Martinez Insua 2008). The development can be seen as an example of the process of grammaticalization, in which *there* has lost (most of) its locative meaning and has been reanalyzed as an empty subject.[21]

The postverbal noun phrase is generally regarded as the 'notional' subject (Quirk et al. 1985; Breivik/Martinez Insua 2008), also called the 'real' or 'logical' subject (e.g. Breivik 1981, 1999; Denison 1999) or 'displaced' subject (Ward et al. 2002; Collins 2012) because it corresponds semantically to the subject of the non-existential counterpart. Compare, for example, the existential sentence in (3.37), *There was nothing wrong*, with its non-existential counterpart, *Nothing was wrong*, in which *nothing* is the subject. The notional subject is normally an indefinite noun phrase, which follows from the fact that the construction mainly serves to introduce (hearer-)new information or inactive referents into the discourse thereby raising it into the addressee's consciousness (Bolinger 1977: 92; Lambrecht 1994: 179; Johansson 1997; Biber et al. 1999: 944; Ward et al. 2002: 1396).

Given these functions, the term 'existential' may be somewhat misleading because the construction does obviously not only assert the existence or non-existence of some entity (although this is one of its main functions). Furthermore, there are other constructions that can be used to express (non-)existence (e.g. *God exists.*). The term should rather be seen as simply describing constructions involving existential *there*

21 There are linguists who claim that the locative meaning is not completely lost in existential *there*. Pfenninger (2009), for example, states that existential *there* has retained some locative colouring from the locative adverb *there*, which "can be understood as a continuation of its original lexical meaning" (2009: 248). She adds that *there* "may function as an indicator of spatial location and thus may contribute in its own right to the meaning of the whole construction" (2009: 249).

and a form of the verb *be*. There are similar constructions with introductory *there* but with a verb other than *be*, for example, *appear*, *arrive*, *come*, *occur*. Example sentences are given in (3.34).

(3.34) a. **There comes** a point where everything you touch becomes cute (ICE-IRE:S1A-058)

b. **There occur** from time to time political events and situations which involve focused and easily discerned spiritual values (ICE-CAN:W1B-025)

These 'presentational' *there*-constructions are rarely found in the ICE corpora and the present study therefore concentrates on existential *there*-constructions only.

As for the information structure of existential clauses, the grammatical subject *there* occupies the topic position which is generally associated with old information, while the notional subject provides the new information. That is, the construction allows for the old-before-new information structure to be preserved, as is nicely expressed by Ward and Birner (2004) in the following words. Note that they group existential clauses within the class of what they call 'postposing constructions'.

> Postposing constructions preserve the old-before-new information-structure pa-radigm by presenting relatively unfamiliar information in postverbal position. That is, when canonical word order would result in the placement of new information in subject position, postposing offers a way of placing it instead toward the end of the clause, in the expected position for new information. (Ward/Birner 2004: 163)

Following Ward et al. (2002), the present study draws a distinction between bare and extended existentials. The former type contains nothing more but the noun phrase in postverbal position or some additional optional adjuncts which have no syntactic significance for the existential construction. Consider the two examples of *there*-existentials in sentence (3.35a). Both examples are bare existentials with the first example containing the optional adjunct *then*. In example (b) the adverbial clause *because the principles have got to be declared* is also of no syntactic relevance for the existential construction and therefore the sentence is classified as a bare existential.

(3.35) a. Then **there was a completely green thing** and it has <,> a round leaf all the way round the stem then it goes up and **there's another round one** (ICE-IRE:S1A-077)

b. **There will be no talking** <,> because the principles have got to be declared (ICE-IRE:S1A-073)

c. And it was <,> the the brake thing. I usen't to drive I used to drive on the main road on Sunday morning. **There was no cars** (ICE-IRE:S1A-079)

d. But **there's big big happenings** (ICE-IRE:S1A-093)

Despite the fact that there are no temporal or locative descriptions in bare existentials, there is an implicit locative, which can often be inferred from the preceding context. In (3.35c), for example, we know from the preceding sentence that it was on the main roads that there were no cars. Bare existentials typically (but not exclusively) predicate the existence of an entity or entities as in (a) to (c) or the occurrence of an event,

as in (d).

In extended existentials, on the other hand, the notional subject is followed by some additional material that is of relevance for the construction. Very common are locative and temporal expansions, illustrated in (3.36) in bold print.

(3.36) a. There's no lift **in my house** (ICE-IRE:S1A-014)

b. There was certain intimacy **in the eighteenth century** (ICE-GB:S1A-020)

The adverbial extensions in these examples locate the new referent, introduced by the notional subject, in space and time. The grammatical subject *there* and the verb *be* contain little information and so we expect the rest of the sentence to be informative in some way. That is, the notional subject and/or the adverbial expansion contribute the informative bits in these existential clauses (Johansson 1997).

Additionally, there are predicative extensions, as in (3.37), infinitival extensions, as in (3.38), participial extensions, as in (3.39), and relative clause extensions, as in (3.41).

(3.37) Thankfully there was nothing **wrong** (ICE-IRE:S1A-059)

According to Ward et al. (2002: 1394), the type of predicative allowed in existential clauses is severely restricted, with only those items denoting temporary states - as opposed to (relatively) permanent properties - being allowed. Nominal predicatives are normally excluded. An exception involves idiomatic expressions with the noun phrase *the matter* (*Is there anything the matter*?).

In existentials with infinitival extensions the notional subject may function either as the subject (3.38a) or the object (3.38b) of the infinitive clause.

(3.38) a. There is huge profit **to be made** (ICE-IRE:S1A-069)

b. There wasn't enough air or something for it <,> **to breathe** (ICE-CAN:S1A-085)

Existential clauses with a participial extension involve an *ing*- or an *ed*-participial clause after the notional subject, illustrated in (3.39).

(3.39) a. There's a festival **starting this week in Derry** (ICE-IRE:S1A-012)

b. There's about twenty people **invited** (ICE-NZ:S1A-039)

The non-existential counterparts of these two sentences would be *A festival is starting this week in Derry* and *About twenty people are invited*. Note that existential clauses with a participial extension are inherently ambiguous and allow for two different readings. First, the participial clauses may be an extension of the existential clause as in the examples in (3.39). Second, the participial clause may be a modifier of the notional subject, that is, part of the noun phrase. Such constructions, exemplified in (3.40), must be distinguished from those in (3.39).

(3.40) a. There's a word **beginning with D** that would describe it (ICE-GB:S1A-018)

b. Did you know there's a a fly **called the Botsi fly**? (ICE-IRE:S1A-015)

The non-existential counterparts of these sentences are not *A word that would describe it is beginning with D* and *A fly is called the Botsi fly*, but the participial clauses belong to the notional subjects: *[A word beginning with D] would describe it* and *[A fly called the Botsi fly] exists*. In these examples we have complex notional subjects but not extended existential constructions.

Finally, there are constructions with a relative clause extension. The examples in (3.41) illustrate the different realizational variants of the relative pronoun, namely *wh*-words, *that* and zero.

(3.41) a. There are some women **who judge a man <,> uh by his car** (ICE-IRE:S1A-061)

b. There's only two **that are nice** (ICE-IRE:S1A-035)

c. There's a wild lot of people **find her okay** (ICE-IRE:S1A-026)

Some linguists suggest that the distinction that can be drawn with participial extensions (the participial is part of the notional subject or an extension) may also apply to existentials containing a relative clause. That is, in addition to relative clause extensions there are items where the relative clause modifies the notional subject. The sentences in (3.41) are examples of relative clause extensions. They are the existential counterparts of the following sentences: *Some women judge a man by his car*; *Only two are nice*; *A wild lot of people find her okay*. By contrast, in the sentence *There's a barrister that was in my class at Queen's* (ICE-IRE:S1A-020) the relative clause is a modifier of *a barrister*. The sentence cannot be paraphrased as *A barrister was in my class at Queen's*. However, in the case of relative clauses the distinction is much less straightforward than in the case of participial clauses and some linguists are rather hesitant to draw such a distinction (cf. e.g. Ward et al. 2002: 1396). Hence, it was decided to make no such distinction in the present study. The two types of existential with participial extensions, however, will be grouped separately in the analysis of existential clauses in Chapter 4.

The notional subject is, as noted earlier, typically an indefinite noun phrase, but definite noun phrases do also occur, provided they represent hearer-new information and/or are uniquely identifiable to the addressee.[22] The instances of such existentials in my data suggest that the speakers using definite noun phrases often assume prior familiarity with the entities. Consider, for example, the existential sentence in (3.42). The conversation is about stomach pain and speaker B says that she usually takes Voltaren. Speaker U then mentions Naprosyn as an alternative. U assumes prior familiarity with this drug and thus the definite noun phrase is admissible.

(3.42) B: When I can feel one coming on what I have to do is chew a piece of bread or something and then eat it as I take the tablet

22 Ward et al. (2002: 1398ff.) present five such cases in which the definite noun phrase satisfies the hearer-new requirement: hearer-old entities are treated as hearer-new; hearer-new tokens are part of hearer-old types; hearer-old entities specify the value of a variable in an open proposition; the content of the noun phrase is sufficient to fully identify the referent; the noun phrase contains 'false definite' *this* ('false definite' because the noun phrase is definite in form but indefinite in meaning). Also see Ward and Birner (1995) for a discussion of these different cases of definite noun phrases.

otherwise it burns the lining of your stomach

U: **There's that Naprosyn**

B: Yeah pretty strong that stuff

(ICE-NZ:S1A-011)

This example could also be understood as a list: Naprosyn and Voltaren are members of the set of drugs. List readings are very often involved in existentials with definite noun phrases. Grzegorek (1984: 154) notes that "[l]ist *there*-sentences do not assert existence. They can be paraphrased as 'one could mention…', i.e. they only bring the referent of the focus NP to the addressee's consciousness". Also consider the following examples. In (3.43) we have a list of Titanic movies and, in addition, the speaker assumes that the addressee is familiar with the nineteen fifties version. In (3.44) the speaker lists the alternatives of 'how you can be in a relationship'.

(3.43) The first movie was <,> uh Titanic in the nineteen thirties and then **there was the nineteen fifties version** (ICE-PHI:S1A-074)

(3.44) And that there are many choices to how you can <,,> be in a relationship
There's the common-law situations there's the more casual situations and there's marriage
And uhm the essential idea of marriage is a life-long commitment
(ICE-CAN:S1A-015)

Existentials with 'false definite' *this*, to borrow Ward et al.'s (2002) term, are also relatively frequent. Illustrative examples are given in (3.45) and (3.46).

(3.45) And uhm <,> I can remember him running back and telling us that the train was coming and then you would start to hear the whistle <,> and then you would start to see the smoke
And it was so gigantic I it was you know when you're a little child and **there's this huge machine coming in**
(ICE-CAN:S1A-042)

(3.46) And um i want to go to New York and Chicago cos **there's this really cool place in Chicago where they filmed a bit of um <,> of Back Draft**
and **there's this really nice view**
and i want to see that view
(ICE-NZ:S1A-043)

In these sentences, demonstrative *this* is used in a sense that is pragmatically equivalent to *a* or *some*, that is, *this huge machine* in (3.45) can be replaced by *a huge machine* but not by *the huge machine*; and *this really nice view* in (3.46) can be replaced by *a really nice view*. Note that in the first example in (3.46), *there's this really cool place in Chicago where they filmed a bit of Back Draft*, demonstrative *this* is admissible for other reasons: in this case the content of the noun phrase is sufficient to fully identify the referent.

Existential clauses are a common feature of conversation which is surprising because they are a specific device used for the introduction of nouns, which, in turn, are generally least common in conversation out of all registers. Biber et al. (1999: 953) suggest that the high frequency in conversation can be explained by the fact that the construction "agrees with the looser syntactic organization of conversation". In this respect the existential *there*-construction is very similar to left dislocation. Both constructions break down a sentence into smaller units and thus make it possible to present one unit of information at a time. Consider, for example, sentence (3.41c) and the corresponding left dislocation construction: *A wild lot of people, they find her okay.*

Finally, note that constructions containing only *there* and *be* are not included in the present study. This may be the case in question tags, as in (3.47), in expanded RDs or *so*-tags, as in (3.48), or assertions and replys, as in (3.49).

(3.47) I was going to say why don't you talk to Laura but there's not much point in that **is there**? (ICE-GB:S1A-038)

(3.48) There's going to be a few sitting and more standing **so there is** (ICE-GB:S1A-072)

(3.49) F: There's a lot to be said for reading your guidebook.
D: **Oh there is**
We were totally innocent
But it's a well-known place actually
(ICE-IRE:S1A-021)

Tokens such as these have been excluded from the present study because they are of not much relevance for the analyses. I am particularly interested in the (non-)concord between verb and notional subject and the distribution of bare and extended existentials across the varieties of English analyzed.

3.2.5 Cleft constructions

This section focusses on the description of the three major types of cleft construction, namely IT-clefts, basic and reversed pseudo-clefts. There are further (sub-)types of cleft construction which are not discussed in the present study. These include inferential clefts (e.g. Declerck 1992; Delahunty 1995; Calude 2009b), *there*-clefts (e.g. Davidse 2000; Lambrecht 2001b), *if-because* clefts and *since*-clefts (e.g. Lambrecht 2001b; Di Tullio 2006).

It-clefts

Prototypical IT-clefts consist of four parts: the introductory pronominal *it*, a form of the copula *be*, a post-copular element (clefted or highlighted element) and a clause

(cleft or relative clause) introduced by *that*, *who*, *which* or zero. They are usually seen as identifying or specificational constructions with the cleft clause being the variable or identified element, and the clefted constituent the value or identifier (e.g. Halliday 1967; Declerck 1984). For example, in *It's Tom who loves apples* we have the open proposition 'x loves apples' and the value of the variable x is *Tom*.

Properties that are usually associated with the IT-cleft construction include an implication of contrast, emphasis and a sense of exhaustiveness. These properties can easily be explained when the construction's features as a specificational or identifying construction are taken into account.[23] Choosing one value for a variable among a set of other options automatically implicates a sense of contrast with the other members of the set. Furthermore, if one value among other alternatives is chosen, it is also naturally emphasized. Exhaustiveness follows from the assumption that a speaker is cooperative and specifies a variable correctly (cf. Grice's Cooperative Principle and conversational maxims; e.g. Grice 1989). If the speaker utters the sentence in (3.7a), *It's Tom that loves apples*, the addressee can conclude that there is only one person who loves apples. If there were more (or less) people who loved apples, then the speaker would deceive the addressee. According to Declerck (1984: 272), it is precisely these properties of clefts - involving a sense of contrastiveness, emphasis and exhaustiveness - that make speakers prefer this type of construction over the non-cleft counterpart in certain communicative situations.

Looking at the realization of the clefted constituent, we typically find a noun phrase functioning as the antecedent of a subject in the cleft clause, as in (3.50a). But there are other functional possibilities, such as direct and prepositional object, adverbial, complement or possessive, as illustrated in the following examples, respectively.

(3.50) a. Anyway it was **Mary Shelley** that wrote that Frankenstein thing (ICE-IRE:S1A-072)

b. It was actually **the study of architecture** I really enjoyed <,,> (ICE-GB:S1A-034)

c. Mm <,> because it's just **the grammar** we're looking at (ICE-IRE:S1A-002)

d. It would've been **the house** they were all reared in (ICE-IRE:S1A-009)

e. They made me secretary, but it wasn't **secretary** I'd wanted to be (Ward et al. 2002: 1418)

f. It's **you** whose head will roll (Ward et al. 2002: 1418)

Most common are examples where the clefted constituent is the antecedent of a subject, object or time adverbial in the cleft clause; complements and possessives are rarely found. Note that the subordinator can be omitted even in cases where it is the subject, although this tends to happen only in more informal contexts.

(3.51) a. It was the authorities were responsible there (ICE-IND:S1A-083)

b. It's Steph Gan Steward and another person went out (ICE-SIN:S1A-058)

23 See, for example, Declerck (1984) and Patten (2012a) for accounts that derive many of the cleft construction's properties from the assumption that it is a member of the family of specificational structures.

The next common type of clefted constituents is prepositional phrases, as exemplified in (3.52). The prepositional phrases can have the form of preposition plus noun phrase (a), preposition plus clause (b), or simply of a preposition (c).

(3.52) a. It must have been just **after Christmas** I met him (ICE-IRE:S1A-020)

b. It wasn't **till I was perhaps twenty-five or thirty** that I read them and enjoyed them <,> (ICE-GB:S1A-013)

c. It's **downstairs** they want to play (Ward et al. 2002: 1418)

Adverbial phrases and adjective phrases are rather rarely found in the position of the clefted constituent. Example sentences are given in (3.53a-c). Finite and non-finite clauses, illustrated in (3.53d-e), are even rarer. Finite clauses are more common in pseudo-clefts.

(3.53) a. So it's **rare** that I get to see them but still the closeness is still there (ICE-PHI:S1A-040)

b. It's only **now** that you're realising a lot of other little things (ICE-NZ:S1A-046)

c. So it's crazy they're you know (ICE-JAM:S1A-014)

d. It is who you know that will get you somewhere (ICE-JAM:S1A-058)

e. It was **listening to Sue's story** that made me realise how lucky we have been (Ward et al. 2002: 1418)

There is also some variation of the initial element and the cleft clause. The former can be realized as a demonstrative pronoun instead of the prototypical *it*, and in place of the cleft clause we also find non-finite clauses. Consider the sentences in (3.54) for illustration.

(3.54) a. **That** was her husband that died (ICE-IRE:S1A-083)

b. **These** are the words I'm introducing today these are words I am introducing tomorrow <,> (ICE-IND:S1A-078)

c. No ironically it would be the guys **telling me** (ICE-PHI:S1A-056)

d. Yeah it's seldom **for Hong Kong people to play such games in Hong Kong** (ICE-HK:S1A-083)

In so-called 'truncated' IT-clefts the cleft clause is omitted. This is possible in cases where its information is recoverable from the context. Consider the examples in (3.55) and (3.56).

(3.55) D: She's very good

E: I'm sure she'll I mean she's not the one I'm worried about **It's you**

(ICE-IRE:S1A-003)

(3.56) A: Who was the one who was picked?
B: Instead of Vince?
A: Yeah
B: **It was Ray Allen of the Milwaukee Bucks**

(ICE-HK:S1A-083)

In these IT-clefts the cleft clause can easily be recovered from the preceding utterances. In (3.55), the full cleft would be *It's you I'm worried about* and in (3.56) it would be *It was Ray Allen who was picked.*

In addition to formal variation, we also find IT-clefts that deviate from the prototypical construction in functional and information-structural terms. Normally, the clefted constituent is the focus representing new information, while the cleft clause represents presupposed or known information.

These constructions are sometimes also referred to as 'stressed-focus' clefts (cf. e.g. Prince 1978) or 'contrastive' clefts (e.g. Declerck 1984) because the main focal accent falls on the clefted constituent, which is thus highlighted, and this type of cleft often serves a contrastive function. The distribution of given and new information is different in so-called 'informative-presupposition IT-clefts' (cf. Prince 1978). In this type of cleft, the cleft clause expresses the new and most important information and often also contains the focus. Consider the following example for illustration.

(3.57) **It was just about 50 years ago that Henry Ford gave us the weekend**. On September 25, 1926, in a somewhat shocking move for that time, he decided to establish a 40-hour work week, giving his employees two days off instead of one.

(Prince 1978: 898)

A speaker uses this type of IT-cleft when he does not expect the addressee to know the information in the cleft clause (*Henry Ford gave us the weekend*). The main function of the construction is to mark the information in the cleft clause as if it was known and "to INFORM the hearer of that very information" (Prince 1978: 898), which he just happens not to be aware of. This property of the cleft has come to be called the *known-fact effect*. Furthermore, this type of cleft may function as backgrounding device in the sense that the information conveyed is "background material [...] subordinate in importance to what follows" (Prince 1978: 902).

In addition to the functional differences, there are also a number of formal differences between the informative-presupposition cleft and the stressed-focus cleft: the subordinators *that* and *wh-* are not deletable, the cleft clause is normally (vs. weakly) stressed, the highlighted element is not as heavily stressed and it is generally short and anaphoric, there is no pseudo-cleft equivalent. While stressed-focus IT-clefts cannot be used as discourse openers, this works well with the informative-presupposition IT-cleft.

Looking at the example in (3.57) it can furthermore be seen that the specificational meaning is less strongly marked in this type of IT-cleft. The cleft *It was just about 50 years ago that Henry Ford gave us the weekend* is hardly felt to specify the value of a variable, an impression which is underlined by the fact that there is no question

word explicitly introducing a variable (Compare the stressed-focus cleft *It's Tom who loves apples*, where the variable is explicitly introduced by *who loves apples*). Given these properties, Declerck (1984: 282)[24] concludes that "[t]he function of this cleft is therefore not the (normally primary) function of identification (i.e. of specifying a value for a variable) but the (normally derived) function of marking emphasis".[25]

In addition to the stressed-focus IT-cleft (the highlighted element expresses new information and the cleft clause old information) and the informative-presupposition IT-cleft (both constituents represent new information), Declerck (1984) distinguishes a third type of cleft where the clefted constituent represents old and the cleft clause new information.[26] Examples of this type of cleft, called 'unstressed-anaphoric-focus clefts' by Declerck, are given in (3.58) (Declerck's examples (25a) and (26a)).

(3.58) a. However, it turns out that there is interesting independent evidence for this rule and **it is to that evidence that we must now turn**

b. But why is everybody so interested in uranium? Because **it is uranium that you need to produce atomic power**

(Declerck 1984: 263)

The clefted constituents *it is to that evidence* and *it is uranium* are anaphoric and thus not heavily stressed. They also invite no contrastive reading. Since there is the antecedent of the clefted constituent in the preceding sentence this type of cleft is not used as a discourse opener.

In sum, as far as the information structure of IT-clefts is concerned, we can distinguish three types: the 'normal' contrastive or stressed-focus clefts (with new information in the clefted constituent and old information in the cleft clause), unstressed-anaphoric-focus clefts (old information in the clefted constituent and new information in the cleft clause) and discontinuous or informative-presupposition clefts (both constituents express new information).

Another construction that resembles IT-clefts and is worth mentioning is called 'predicational' or 'proverbial' IT-cleft (e.g. Prince 1978; Declerck 1983; Lambrecht 2001b; Ward et al. 2002), illustrated in (3.59).

(3.59) a. It's a wise child that knows its own father

b. It's a fortunate man that will find the fountain of youth

(Prince 1978: 905)

Note that these sentences do not have the same meaning as the basic sentence equivalents *A wise child knows its own father* and *A fortunate man will find the fountain of youth*. The meanings are rather *A child that knows its own father is a wise one* and *A man that finds the fountain of youth is fortunate*. The copula *be* is here used in its predicational and not in its specificational sense. Given these differences between

24 The stressed-focus IT-cleft is called 'contrastive cleft' by Declerck and the informative-presupposition IT-cleft is called 'discontinuous cleft'.

25 For further discussions of this type of IT-cleft also see Geluykens 1988, Delin/Oberlander 1995, Lambrecht 2001, Ward et al. 2002, Patten 2012.

26 Prince (1978) groups these sentences into the class of informative-presupposition clefts.

the IT-clefts and the basic sentences, some authors see the predicational cleft as not belonging to the group of IT-clefts, for example, Lambrecht (2001: 503) who suggests a right-dislocation interpretation instead. Others, however, see these constructions as IT-clefts (e.g. Prince 1978; Patten 2012). No such examples have been found in the database of the present study.

Pseudo-clefts

Prototypical pseudo-clefts consist of three parts. In the basic form, illustrated in (3.60a), this includes an initial fused relative clause (also called cleft clause, *What Tom loves*)[27], a form of the copula *be* and a highlighted element as the complement of *be*, here *apples*. In the reversed pseudo-cleft, illustrated in example (3.60b), the order of these three constituents is inverted.

(3.60) a. What Tom loves is apples.

b. Apples is what Tom loves.

Pseudo-clefts are also identifying copular constructions with the cleft clause defining a variable whose value is specified by the highlighted element. That is, in (3.60) the cleft clause corresponds to 'the x [Tom loves x]' and the highlighted element 'apples' is the value of the variable x.

There is a range of different elements that can function as the clefted constituent in pseudo-clefts. The example in (3.60a) represents the common type of a noun phrase functioning as the highlighted element. Also quite common are finite or non-finite clauses. The sentences in (3.61) illustrate a declarative finite clause, (a) and (b); an interrogative clause, (c); a *to*-infinitive, (d); a bare infinitive, (e); and an *ing*-participle, (f).

(3.61) a. What happened was **that they had to flee India** (ICE-JAM:S1A-041)

b. **He lost a nephew** was basically what happened uh from suicide (ICE-CAN:S1A-016)

c. What we look at is **how languages behave on a societal level** (ICE-JAM:S1A-044)

d. What I've always tended to do is **to do my own stretches at home** (ICE-GB:S1A-003)

e. What you can do is **put three in a chain** (ICE-GB:S1A-067)

f. What we're trying to do is **trying to stay away from myth and legend** (ICE-CAN:S1A-039)

27 The term 'fused relative' is used by Ward et al. (2002) and Collins (1991). Other terms that are found in the literature include 'headless' relative (Oberlander/Delin 1996), 'nominal relative clause' (Quirk et al. 1985) and 'free relative' (widely used amongst transformational linguists).

Clauses can more easily be integrated into pseudo-clefts than in IT-clefts. As can be seen in sentences (d) to (f), pseudo-clefts allow for the focus to fall on the verb by using the substitute verb *do*. This is normally not admissible with IT-clefts (or it definitely sounds rather unnatural): **It's to do my own stretches at home that I've always tended to do*; **It's put three in a chain that you can do*; **It's trying to stay away from myth and legend that we're trying to do*.[28]

Adjective phrases and prepositional phrases are rather rare in pseudo-clefts. The latter type is predominantly found in IT-clefts.

(3.62) a. **Insensitive** is how I'd be inclined to describe him

b. **In the morning** is when I do my best work

(Ward et al. 2002: 1422)

In addition to the clefted constituent, there is also variation of the initial element of the cleft clause, yielding so-called ALL-clefts as in (3.63) and TH-clefts as in (3.64). In (3.63), speaker A, who has to move out of her flat soon, complains about having to do all the cleaning while her flat-mate is out.[29]

(3.63) B: What about Maureen
Is she out

A: She's out
The fucker little [several syllables unclear]

B: And did she clean anything else
Suppose she didn't really live there

A: No well I would only've I'd expected her to do her room anyway
All I need to do is run the hoover over the place clear up my stuff and dust and <,> scrub the kitchen floor and the bathroom
You know what I mean
I'm obviously I'll clean the oven

B: That's it
That's all you need to do

(ICE-IRE:S1A-095)

In ALL-clefts we find *all* in place of the more prototypical *what*. Structurally, the constructions are very similar; note that there is also a reversed version of ALL-cleft (cf. B's third turn). Semantically, however, there is some difference between the two types of pseudo-cleft. Compare the ALL-cleft in A's second turn, *All I need to do is run the hoover over the place*, to its WH-cleft counterpart, *What I need to do is run the hoover over the place*. The meaning of the WH-cleft is that I need to do something and this something is hoovering and nothing more. The ALL-cleft, on the other hand,

28 See Mair and Winkle (2012) for a detailed discussion of the *do*-type of pseudo-cleft.

29 Note that speaker A's *all*-cleft is followed by a reversed *all*-cleft uttered by speaker B later in the discourse. Calude and Miller (2009) show that clefting is contagious in spontaneous spoken interaction, that is, speakers emulate each other's grammar and use the same constructions. The extract in (3.63) supports their claim.

conveys the meaning that there is not very much I need to do; it is only the hoovering that needs to be done and nothing more. In the reversed version, there is also a semantic difference. Reversed pseudo-clefts often serve a 'summative' function, that is, they often occur at the end of a discourse, summing up what has been said before (e.g. Collins 1991; Miller/Weinert 1998). Both the reversed ALL-cleft in (3.63) and its *wh*-cleft equivalent *That's what you need to do* serve this 'summary' function, but the reversed ALL-cleft has an additional sense of 'Don't worry, you don't have to do more than that'.

In TH-clefts, the initial elements are pro-nouns such as *person, one, place, time, reason, way* instead of *what*.[30] They also exist in the basic variant, as in (3.64a-c), and in the reversed version (3.64d-e).

(3.64) a. **The only person** who might have his own secretary is the director you know (ICE-JAM:S1A-027)

b. **The only area** that we need to have some improvement in is tourism (ICE-JAM:S1A-049)

c. Yeah but **the main reason** why I didn't want to go there is because of the financial <,> (ICE-HK:S1A-008)

d. John's **the one** that does the training (ICE-GB:S1A-028)

e. You're **the only person** who was there (ICE-GB:S1A-064)

These pro-nouns provide one way of compensating for the inadmissibility of WH-words other than *what* in pseudo-clefts, for example, *who* as in **Who that did was my brother* (Halliday 1967: 234; Quirk et al. 1985: 1388-9). Note that the pro-noun is always introduced by the definite article *the*.

The class of reversed pseudo-clefts comprises an additional type that is sometimes seen as a type of cleft on its own. This type of reversed pseudo-cleft involves an initial demonstrative pronoun *this* or *that* and is thus also referred to as 'demonstrative cleft' (e.g. Biber et al. 1999; Calude 2008). Examples of this type of cleft have been mentioned before, *That's what he loves* in (3.65) and *That's all you need to do* in (3.63). Due to the demonstrative pronouns in initial position demonstrative clefts can have anaphoric, cataphoric or exophoric ties to the surrounding discourse, that is, the clefted constituents direct the addressee to the value of the variable (i.e. the open proposition in the cleft clause) without actually naming it. Rather, it can be found in the preceding discourse, in the immediately following discourse or it points to the physical environment (if the demonstrative pronoun is used as a spatial deictic).

In contrast to the other types of cleft, the major discourse function of demonstrative clefts is not that of highlighting the clefted constituent (although it can be used for that purpose) but rather to organize the information. They can be used to provide explanations (*That's what it sounds like*; ICE-GB:S1A-085), to express the speaker's

30 Note that there are actually two classes of items that occur in initial position in TH-clefts, namely 'pro-nouns' and 'general nouns'. While Collins (1991: 31) argues that TH-clefts headed by general nouns should not be accepted as pseudo-clefts because this would lead to the "open-endedness of the class", other studies do not make such a distinction and include all types in the class of pseudo-clefts (also see Halliday 1967: 233f.). The present study follows Collins (1991) and includes only the restricted set of pro-nouns given above.

opinion or evaluation of entities mentioned in the preceding discourse (*That's how I think*; ICE-PHI:S1A-081), to express agreement with something that has been said before (*Exactly that's what I'm saying*; ICE-JAM:S1A-009). What all these various roles of the demonstrative cleft have in common is that they organize the discourse.[31]

A number of studies have found that demonstrative clefts are much more frequent in (spontaneous) spoken interaction than the other major types of clefts discussed above (e.g. Collins 1991; Miller/Weinert 1998; Biber et al. 1999; Calude 2008). This is mainly due to the make-up of the construction and its major function of managing the information flow. They are ideally suited to spoken interaction as they are low in information content and cognitive load which allows for ease of production and processing on both parts, the speaker and the addressee. Furthermore, they are described to have a rather informal nature (Biber et al. 1999).

Regarding the information structure of pseudo-clefts, Declerck (1984) distinguishes between the same three variants as with IT-clefts.[32] In contrastive pseudo-clefts, the cleft clause expresses (discourse-)old or presupposed information while the highlighted element represents new information. The latter is heavily stressed and therefore contrastive. The unstressed-anaphoric-focus type of pseudo-cleft normally occurs only in the reversed version. In this type of cleft the cleft clause represents new information (but represented as if it was old) and the highlighted element represents old information. Consider the following example of this type of cleft.

(3.65) A: Why does Tom buy so many apples?
B: Because that's what he loves.
*Because what he loves is that.

Pseudo-clefts can also be used as discourse openers presenting new information in both constituents, cleft clause and clefted constituent (the informative-presupposition or 'discontinuous' type). Consider, for example, the sentence in (3.66) which might be the opening sentence in a lecture.

(3.66) What we will talk about today is the transcendence of pi.

Since this is the first sentence of the lecture the cleft clause can obviously not be old and derived from the preceding discourse. The information is new, but it is represented as if it was known. Put differently, a speaker opening his lecture with this sentence can assume that the addressees know that he has intentions about what he will discuss in the lecture, that is, the cleft clause *What we will talk about today* does not really come as a surprise.[33] A similar situation obtains in the case of clefts where the cleft clause expresses highly general information, for example, *what I mean* or *what I want to say*. Pseudo-clefts with such meta-linguistic cleft clauses are very common and, according to Patten (2012a: 242), have "developed a presentational function, in which the act of specification is merely used as a device for introducing a focal

31 For a detailed analysis of the demonstrative cleft's discourse functions see Calude (2008).
32 See Oberlander and Delin (1996) for criticism of this differentiation.
33 Also see Ward et al. (2002: 1426), who suggest that the (discourse-)old status of the information in the cleft clause "may derive from the context of the situation rather than from prior mention".

proposition". That is, we have the same situation as with informative-presupposition IT-clefts, which are used for the marking of emphasis rather than for specifying the value of a variable.

3.3 Previous research

This section reviews research that has been conducted on dislocation, fronting, existential *there*-constructions and cleft sentences. The literature review does not aim at being exhaustive but will only include those (aspects of) works which are relevant to the present study. It will review discussions on basic conceptual and terminological issues as well as cross-varietal and acquisitional research.

Dislocation constructions

Left and right dislocation constructions have been extensively discussed in the literature where they are given various labels, for example, Theme and Tail (Aijmer 1989), head and tail (e.g. McCarthy/Carter 1997; Rühlemann 2006; Timmis 2010), anticipated and postponed identification (Quirk et al. 1985: 1310) or amplificatory tag (Quirk et al. 1985: 1417), preface and noun phrase tag (Biber et al. 1999: 957), reference and substitution (Halliday 1967). Some authors deliberately chose not to use the term 'dislocation' and challenged its appropriacy. For them, the term carries negative connotations or is seen as a misnomer or a misleading metaphor (cf. e.g. McCarthy/Carter 1997; Rühlemann 2006; Timmis 2010). Given that dislocation constructions are almost exclusively conversational phenomena, these authors argue that the term 'dislocation' implies that spoken grammar is marked, insufficient or erroneous in comparison to written grammar, which is taken as the unmarked norm. McCarthy and Carter (1997: 407), for example, object to the term on the ground that it implies that the structure is "some kind of aberrant variation on a 'normal' structure", and thus prefer the term 'tails' for 'right dislocation' and 'headers' for 'left dislocation' (Carter/McCarthy 2006). Rühlemann (2006: 394) argues that "the term 'dislocation' covertly evaluates heads and tails as syntactic choices that fail to meet the standards of the written language and, as a result, devalues them". He, then, also goes with the terms 'head' and 'tail'.

I agree with these authors that spoken language should in no way be seen as aberrant, dysfunctional, erroneous or inferior to written language. It is just another mode of discourse with its very own preferred structures which are well suited to the real-time and interactional nature of talk (as opposed to the more composed nature of writing). Yet, I stick with the terms 'left dislocation' and 'right dislocation' since for me these terms carry no negative connotations. And since my analysis is not based on the assumption that dislocation constructions involve movement of a constituent

from a basic (canonical) to a derived (dislocated) position, the terms 'left dislocation' and 'right dislocation' are simply labels with no deeper metaphorical meanings.

Alongside these more theoretical and functional discussions, dislocation constructions have also been looked at from a sociopragmatic and cross-varietal perspective in recent years. Durham (2011), for example, analyzes right dislocation structures (RDs) in a corpus of Yorkshire English. She distinguishes three types of RD, namely standard right dislocation (SRD), as in (3.67a), expanded right dislocation (ERD), in which an operator is included within the dislocated element, as in (3.67b), and reverse right dislocation (RRD), in which operator and noun phrase are inverted in the dislocated element, as in (3.67c).

(3.67) a. I was a little angel **me**. [SRD]

b. He stayed with this other woman **John did**. [ERD]

c. She got a great bargain **did her Mum**. [RRD]

(Durham 2011: 261)

Reviewing literature on the history of these three types of RD, Durham (2011: 262) notes that the use of SRDs can be traced back to Old English (quoting Visser 1963: 54) while the two other variants are more recent phenomena dating back to the 19th century (quoting Poutsma 1928: 172; Jespersen 1949: 67). Among the two forms with operators, RRDs are the more common variant and while they are today regarded as idiosyncratic dialectal features of Northern English they are "not viewed as particularly regionally restricted from a historical perspective" (Durham 2011: 262).

Comparing the use of the three types of RD in the speech of men and women, and in the speech of young speakers (aged 15-35), middle-aged speakers (aged 36-69) and older speakers (aged 70+), Durham (2011: 267-8) finds no statistically significant differences. The examination of gender and age together, however, reveals an interesting pattern: RDs are most frequently used by the oldest speakers in both the female and the male data. However, while for the female speakers the frequency of RDs continuously decreases from the older to the younger speakers, there is a v-shaped distribution for the male speakers, with the younger speakers using increasingly more RDs than the middle-aged speakers. This unexpected distribution in the male data is explained by Durham (2011: 273) as an instance of "recycling", a term which she adopts from Dubvois and Horvath (2000) and which describes the situation when "traditionally local features are regaining ground in the youngest generation (particularly young men)" (Durham 2011: 273) resulting in a v-shaped pattern of distribution. As RDs (and in particular RRDs) are used so unexpectedly frequently by the young male speakers in her data, Durham suggests that these forms signal an association "with Northern identity for the young men in York" and that "their increased use is tied to that" (2011: 276).

The expanded form of RD has also been identified as a feature of South African Indian English (SAIE). Mesthrie (1992: 115) argues that "topicalisation in SAIE goes well beyond that of mainstream English varieties, in terms of both syntax and pragmatics", underlining his point by giving six arguments, including declarative tags (e.g. *We paid seventy-six cents we paid* and *We stayed in the Finn Barracks we stayed*).[34]

34 For Mesthrie, topicalization comprises fronting constructions and dislocation constructions.

Given these findings, it is worth analyzing the expanded form of RD across varieties of English. As noted earlier (cf. section 3.2.2), a specific type of RD token has been identified in the Irish English data - the *so*-tag - which is quite similar to Durham's ERD (*That's pure selfish so it is.*).

Another very recent investigation of dislocation structures is provided by Lange (2012) who examines the syntax of spoken Indian English and thus adds a cross-varietal perspective to the discussion of the construction. Comparing the use of dislocation in the Indian and British components of the International Corpus of English (ICE-India and ICE-Great Britain, respectively), Lange (2012: 160) finds that left dislocation constructions (LDs) are overwhelmingly more frequent in the Indian English data than in the British English data (14.18 and 1.56 instances per 10,000 words, respectively), while RDs occur very rarely in ICE-India (0.64) and are more frequent in ICE-GB (2.43).

Searching for possible explanations for the high frequency of LDs in ICE-India, Lange suggests that for many speakers the main motivation may be topic continuity. They form 'identity links' by repeating the salient noun phrase of the immediately preceding utterance (2012: 165). This function, which Lange (2012: 166-7) claims "has so far not been described for LD constructions in other varieties of English", seems to be a particular feature of Indian English as it also applies to other focusing strategies under consideration in her analysis (cf. existential *there*-constructions and fronting constructions below).

In addition to these sociopragmatic and cross-varietal studies, dislocation constructions have also been discussed by researchers in the fields of (first and second) language acquisition and language teaching, where it has been found that the construction is a common feature of learner discourse (cf. e.g. Carter/McCarthy 1995; Williams 1987; Ortega 2009). Carter and McCarthy, for example, argue that in LD tokens there is a 'slot' available before the core constituents of the clause are realized and that speakers choose to fill this slot "for textual and interpersonal motives" (1995: 149f.). Since the items that are put into the available slot are grammatically indeterminate, they state, it is quite easy for language learners to manipulate this slot (Carter/McCarthy 1995: 150). This may explain language learners' tendency to use dislocation constructions.

Fronting constructions

Fronting has been described in the literature as a common feature of New Englishes (e.g. Mesthrie 1992; Alsagoff/Lick 1998). Lange (2012), in her comparison of Indian English with British English, finds that fronting constructions are strikingly more frequent in ICE-India than in ICE-Great Britain (187 vs. 19 tokens). She suggests that this finding may be accounted for by looking at substrate influence as the decisive factor and gives two reasons: firstly, word order constraints in Indian English are possibly more flexible due to influence from various Indian languages which "allow movement of more or less any constituent to sentence-initial position" (2012: 151). This means that speakers of Indian English map the Indian fronting construction onto English and

are thus more willing to use it in general. Secondly, the range of discourse functions is extended by the function of 'elliptical repetition' or 'identity link', which refers to the repetition of discourse elements from the immediately preceding discourse (2012: 152), a feature of Indian English which she has also identified for left dislocation constructions (cf. above). According to Lange, this feature of fronting constructions can be regarded as an "instantiation of the pan-Indian 'grammar of culture'", since it can also be observed with other constructions and can be traced back to substrate influence (2012: 152).

Alongside the New Englishes, fronting constructions have also been claimed to be particularly frequent in the 'Celtic Englishes' (Filppula 1999, 2009; Hickey 2005, 2007). Accounting for this observation in Irish English, Filppula (1999) gives evidence that both the superstrate and the substrate may have contributed to this outcome, possibly also non-contact-influenced factors. He argues that the higher usage of fronting in Irish English may be due to diffusion from earlier stages of English since it can still be found in conservative regional British English dialects (1999: 266). Furthermore, he states that parallels in the Irish language and high occurrences in Hebridean English and Welsh English can be seen as evidence of substrate influence. As far as language universals are concerned, it has to be noted that fronting constructions are widely used among the languages of the world as means of achieving prominence. Undecided about the role of each influencing factor, Filppula concludes that "multiple causation remains the safest conclusion" (1999: 270).

Existential *there*-constructions

The type of construction dealt with in this section was coined existential sentence by Otto Jespersen back in 1924 (1924: 155). Since then a huge body of literature has been produced on the construction, including theoretical accounts and empirical studies in book-length (e.g. Milsark 1979; Hannay 1985; Breivik 1990; McNally 1997; Pfenninger 2009) or in the form of articles published in journals (e.g. Hay/Schreier 2004; Breivik/Martinez-Insua 2008; Collins 2012) or collections (e.g. Grzegorek 1984; Johansson 1997; Breivik 2000). These studies discuss and present different perspectives on issues such as the historical development of existential *there*[35] from the locative adverb *there*, the status of *be* (auxiliary vs. copular), the status of existential *there* (locative vs. non-locative vs. something in between), the nature of complex *there*-constructions, and variable concord (e.g. *There's four books* vs. *There're four books*).

In what follows, I will concentrate on studies that offer a cross-varietal perspective on the construction. Lange (2012) compares existential *there*-constructions in Indian English and British English and notes that overall there seems to be not much difference in the nature and usage of the construction in these two varieties, with

35 The grammatical subject *there* in existential clauses is often called 'existential' *there*, sometimes also 'expletive' *there*. These labels are used to distinguish existential *there* from the locative adverb *there*. The former term carries no explicit meaning (i.e. that *there* may express existence itself) but is a convenient way to refer to the *there* in existential clauses. The term 'expletive' *there* is used by linguists who assume that *there* has lost all its meaning during the process of grammaticalization.

the exception of extended existentials, which are significantly less frequent in Indian English. Lange points to another type of existential, the non-initial existential *there*-construction (e.g. *Night ten o'clock a bus is there <,> but it is <,,> inter-state bus*; ICE-IND:S1A-012), which she claims is exclusive to Indian English (2012: 94). In contrast to 'canonical' *there*-existentials, which mainly serve to introduce hearer-new referents, non-initial *there*-existentials are mainly used to assert the existence or occurrence of an entity that can be retrieved from the preceding discourse, that is, the 'notional' subject refers to hearer-old or even discourse-old entities. Giving a number of examples from Dravidian and Indo-Aryan languages, Lange argues that substrate influence may well account for the construction in Indian English, suggesting that "[t]he 'export' of the pan-Indian construction 'focus NP + be' to become 'focus NP + be + there' in IndE then keeps the structure of the target language English intact, by enlisting nonreferential *there* to fill the slot, so to say" (2012: 119).[36]

There is a large number of studies on variable concord in existentials, that is, the variation between plural concord and singular concord in existentials with plural notional subjects (*there's four books* vs. *there're four books*).[37] Variable concord has been studied in different varieties of English, including British English (Martinez Insua/Palacio Martinez 2003; Crawford 2005; Rupp 2005), American English (Schilling-Estes/Wolfram 1994), Canadian English (Meechan/Foley 1994; Walker 2007), Australian English (Eisikovits 1991) and New Zealand English (Britain/Sudbury 2002; Hay/Schreier 2004). More recently it has been investigated in a number of ICE corpora (Jantos 2009; Collins 2012). These studies on variable concord are difficult to compare because they are based on very different types of data and the constructions are defined differently. However, some common effects can still be noted. There is evidence that variable agreement is in part determined by the grammar and by extralinguistic factors. For example, singular concord is favoured in speech over writing and seems to be more frequent in more informal spoken material; age, sex and level of education seem to play a role, with younger, non-professional males using singular agreement more often than speakers of other profiles. As for grammatical factors, it has been noted that singular concord is favoured in contexts where *be* is in the present tense and is contracted rather than in the past tense and non-contracted. A caveat is in order here: the factors of tense and contractedness are interrelated in the sense that contracted forms almost exclusively occur in present tense contexts. That is why Walker (2007) considers only non-contracted forms. He finds past tense forms to covary with singular concord. More on this in section 4.4.

Different hypotheses are suggested in the literature on the occurrence of singular concord in existential *there*-constructions. One view holds that singular concord re-

36 The non-initial existential *there*-construction indeed seems to be a characteristic feature of Indian English because hardly any instances can be attested in the other corpora analyzed in the present study. Note, however, that the construction is not unique to Indian English as the following examples from Jamaican English show: *I believe that women have the opportunity to achieve anything they want to once they set their mind to it and* ***obstacles will be there*** *yes but <,> as long as you're determined to achieve something nothing will really stop you <,>* (ICE-JAM:S1A-085); ***The pressure was there*** *but <,> we pulled through* (ICE-JAM:S1A-059); ***Silk is there*** *from <,,> uh thirties or forties or something* (ICE-JAM:S1A-006).

37 Instead of the term 'concord' some authors use the term 'agreement'. Definitions of 'concord' can be found in Quirk et al. (1985: 755), Biber et al. (1999) and Huddleston/Pullum (2002).

flects lexicalization (e.g. Hannay 1985; Meechan/Foley 1994; Schilling-Estes/Wolfram 1994; Crawford 2005). That is, it is assumed that the form *there's* has become a single lexical unit which is used in existentials regardless of the number of the subject (compare French *il y a* and Spanish *hay*). Other hypotheses state that variable concord is due to stylistic or social reasons with singular and plural concord being "associated with different degrees of formality" (Schütze 1999: 478). Hay and Schreier (2004: 219) suggest to account for the high frequency of the form *there's* in terms of processing load. They argue that the form *there's* has an "articulatory advantage" over the form *there is* because the former consists of only one syllable.

Hay and Schreier (2004), in their study on variable concord in New Zealand English, note apparent-time effects, that is, older speakers use singular concord less frequently than younger speakers. Given this finding it is plausible to assume that singular concord is on the rise, a "by-product of 'colloquialization' of late Modern English", as Collins (2012: 60) puts it.[38] Further evidence in favour of the colloquialization hypothesis involves the strong association of singular concord with contraction and its preference in speech over writing.

Collins (2012) investigates singular concord in nine ICE corpora (only the direct conversation files) and finds that the L1 English varieties (or Inner Circle varieties, as he calls them) are far more accepting of singular concord than the L2 English varieties (or Outer Circle varieties), with Australian English showing the highest proportion (52.7%), followed by American English (44.4%) and then British English (25.4%). Among the L2 English varieties, Hong Kong English speakers use singular concord most frequently (20.7%), followed by Singapore English (19.6%), Philippine English (17.1%), Indian English (7.0%) and Kenyan English (5.8%). Examining a number of the features which have been identified in the literature as influencing singular concord, Collins finds a strong correlation of singular concord with present tense across all varieties he analyzes. The correlation with contraction is strong in all varieties but strongest in the L1 varieties. The correlations of singular concord with bare versus extended existentials and plural marking of the noun phrase is less significant.

Jantos' (2009) study investigates variable concord in four ICE corpora, namely the complete spoken sections of the British, Jamaican, Indian and Singaporean components, and in the Santa Barbara Corpus of Spoken American English, including a comparison of register variation in the former group of corpora. She is interested in (non-)concord in general and thus includes in her study both *there*-existentials with singular notional subjects and those with plural notional subjects.[39] Jantos finds overall rather low rates of non-concord, with the proportion being highest in American English (17.4%), followed by British English (7.0%) and Singapore English (6.9%).[40] Indian English (5.7%) and Jamaican English (4.9%) show the smallest proportions. As for the different linguistic factors reported in the literature to have an impact on (non-

38 Colloquialization refers to the tendency of written norms to become more informal and move closer to speech (cf. Mair 1997).

39 Collins (2012) and the present study examine only those existentials that contain plural notional subjects. Non-concord in existentials with singular notional subjects (e.g. *There were a referee*; ICE-IND:S1A-080) is very rare. Also see Martinez Insua/Palacios Martinez (2003) who note that "almost all concord variability occurs with plural NPs" (2003: 264).

40 The much higher percentage of non-concord in American English, Jantos (2009: 149) suggests, is probably due to the different make-up of the corpus.

)concord, Jantos finds a significant influence of tense only in British English (present tense covaries with non-concord) (2003: 154ff.). Contracted forms are strongly associated with non-concord in British English, American English and Singapore English but not in Indian English and Jamaican English (2003: 156ff.). With respect to polarity, a tendency of non-concord to occur in sentences with positive polarity can be attested for all varieties analyzed, but the findings are not statistically significant (2003: 159ff.). As far as register variation is concerned, Jantos finds the highest frequency of non-concord in the more informal texts.

Finally, existential *there*-constructions with relative clause extensions have been given some mention in the literature. Particularly interesting is the issue of zero relatives in subject position (e.g. *There's a wild lot of people find her okay*). In present-day English, relative clauses allow zero marking only in object function and subject-zero relatives are usually acceptable only in existential *there*-constructions and *it*-clefts (Fischer 1992: 307).[41] Subject-zero relatives are predominantly found in colloquial spoken English, but they are reported to be widespread "in colloquial English among speakers of various geographical and social backgrounds" (Lodge 1979: 169).[42] For example, Lodge (1979) documents the use of subject-zero relatives in existential *there*-constructions in a British dialect (spoken in Stockport, near Manchester). Tottie and Rey (1997), studying earlier African American Vernacular English, claim that the frequent omission of subject relative pronouns "underlines the character of EAAVE as an English vernacular" (1997: 244). Christie (1996) examines Jamaican relative clauses and notes that subject-zero relatives are common in *there*-existentials. Preusler (1938, 1942) suggests that Irish is a possible source for the establishment of zero relatives or contact clauses, as he calls them, in English. In contrast, Filppula (1999) is more doubtful about the possibility of Irish as a source for the construction and states that "its influence must be secondary only, considering the widespread nature of the same feature in other varieties" (1999: 185).

Zero-subject relatives are also described as a common feature of the speech of Hong Kong English speakers (e.g. Newbrook 1988, 1998; Li 2000; Hung 2012). Newbrook (1988) notes that the omission of subject relative pronouns is very common, much more common than in Singapore English, especially in sentences with existential *there* or *it* and the verb *be* (1988: 30). Interestingly, Newbrook regards the omission of subject relatives as an error, "the most common error made by Hong Kong students in using the relative clause" (ibid.). As possible explanations for the high incidence of subject-zero relatives he suggests that "the error is so common locally that most students would almost automatically produce it" (1988: 31) and adds that

41 In contrast to present-day English, zero marking was frequently found in both subject and object position in Old English and Middle English relative clauses (Fischer 1992: 307). The original relative marker in English was *that*, which developed out of a demonstrative pronoun. The *wh*-words only began to be used as relative pronouns in Middle English. The more formal character attached to them today is probably due to the fact that they arose as "change from above" and when they were first used they were confined to formal contexts (Tagliamonte 2006: 493).

42 Zero-relativization in subject position is also included in the catalogue of features of non-standard varieties of English in the *Handbook of varieties of English* (Kortmann/Schneider 2004). Contrary to the literature reviewed below, it is described to be very rare among L2 English varieties (Kortmann/Szmrecsanyi 2004: 1199). According to the *Handbook*, it can most frequently be attested for L1 varieties (ibid.).

most students have probably never been taught that subject relatives can never be omitted.

In the case of existential *there*-constructions, some researchers take another position and do not regard the subject-zero relatives as errors (e.g. Schachter/Celce-Murcia 1977; Ortega 2009). Finding many sentences of the type *There are so many Taiwan people live around the lake* in the writings of Chinese learners of English they suggest that the students simply try to structure their sentences according to the principles of their topic-prominent L1, that is, that "these errors can also be viewed as attempts to establish a topic and follow it with a comment, a process that is syntactically and pragmatically acceptable in Chinese" (Schachter/Celce-Murcia 1977: 445-446). Ortega (2009) takes up the idea that speakers tend to map the information structure of their L1 onto English sentences and suggests that existentials with subject-zero relatives are part of a developmental continuum from most L1-like to most L2-like information structure. Interestingly, left dislocation constructions constitute the most L1-like information structure in Ortega's continuum, illustrated in (3.68).

(3.68) a. Many Taiwan people, they live around the lake.

b. There are so many Taiwan people live around the lake.

c. There are many Taiwanese people who live around the lake

d. Many Taiwanese people live around the lake.

(borrowed from Ortega 2009: 45)

This continuum is based on the assumption that beginning learners of English tend to frequently map the information structure of their L1 onto English clauses and therefore use left dislocation constructions more often than native English speakers. When they get more familiar with the English language they realize that in this subject-prominent language new information can well be introduced by means of the existential *there*-construction and "[w]ith time and increasing proficiency, the tendency to transfer the information structure of the L1 in order to frame ideas in the L2 may gradually diminish, but the process may be rather slow" (Ortega 2009: 46).

These hypotheses will be tested in the present study, concentrating on the questions of whether left dislocation and existential clauses are indeed used more frequently by L1 Chinese learners of English and whether there is some correlation between these two types of construction as suggested by the developmental continuum. If this is the case, it can be expected that those speakers who are less advanced in their proficiency in English use left dislocation and *there*-existentials of the type given in (3.68b) more frequently than more advanced L2 English speakers. More precisely speaking, different frequencies of use can be expected for Singapore English speakers as compared to Hong Kong English speakers. Complementing the analysis with data from Jamaican English, Indian English and Philippine English, the present study furthermore addresses the question of whether the developmental continuum presented in (3.68) might be universal, that is, applicable to learners of English in general with the typology of the background language(s) playing no prominent role.

Cleft constructions

An extensive body of research has been carried out on cleft constructions over the last decades. Numerous articles have been published in journals or edited volumes, addressing such issues as the cleft's historical origins or derivation (e.g. Akmajian 1970; Gundel 1977; Ball 1994; Filppula 2009; Patten 2012b), their structural properties (e.g. Geluykens 1988; Delin/Oberlander 1995), their semantics and functions in discourse (e.g. Prince 1978; Declerck 1984; Herriman 2004), their presence across various mediums and genres (e.g. Weinert/Miller 1996; Di Tullio 2006), their prosody and information structure (e.g. Collins 2006; Lambrecht 2001b) and their distribution across varieties of English (e.g. Beal 2012; Mair/Winkle 2012; Lange 2012). Additionally, there are book-length analyses available (e.g. Collins 1991; Calude 2009a; Patten 2012a; Reeve 2012).

The following paragraphs will review a number of controversial issues that are debated in the literature. Furthermore, the main aspects of the cross-varietal studies of cleft constructions will be summarized.

As has been seen in the description of IT-clefts in section 3.2.5, the structure, which consists of four parts, is rather complex and difficult to make sense of. Researchers interested in the make-up and function of the construction disagree, for example, on the role of the pronoun *it* and the interpretation of the cleft clause. The explanations depend to a large extent on how the IT-cleft construction is seen in relation to other constructions in the language. Expletive approaches see IT-clefts in relation to the corresponding canonical sentence (*It is Tom who loves apples* vs. *Tom loves apples*). It is assumed that the two sentences are truth-conditionally equivalent and that the cleft is mainly used as a device for focus marking (e.g. Collins 1991; Haugland 1993; Ward et al. 2002). Extraposition approaches, on the other hand, relate the construction to other specifying copular constructions (e.g. *The one who loves apples is Tom*). In the former approach the initial pronoun *it* is seen as the dummy subject which provides no semantic contribution to the cleft construction, whereas researchers following an extraposition approach ascribe more meaningful roles to it (yet, they do not agree on this role; cf. e.g. Akmajian 1970; Patten 2012a; Reeve 2012).

As for the cleft clause, it superficially looks like a restrictive relative clause, but problems with such an understanding arise, for example, in sentences such as the one in (3.7a), *It's Tom who loves apples*. The antecedent of the relative clause is the proper noun *Tom*, which cannot normally be modified by a restrictive relative clause. It has therefore been suggested to take the cleft clause as a complement of the copular *be* (e.g. Sornicola 1988). Among those who regard the cleft clause as a restrictive relative clause, there is disagreement on whether it modifies the clefted constituent (e.g. Lambrecht 2001b; Ward et al. 2002) or the initial pronoun *it* (cf. Patten 2012a).

More recently, IT-clefts have also been studied from the perspective of construction grammar (e.g. Davidse 2000; Lambrecht 2001b; Patten 2012a). In these approaches cleft constructions are seen as symbolic pairings of form and meaning. It is assumed that the construction has idiosyncratic properties that cannot be explained on the basis of general grammatical rules, that is, they cannot be accounted for in terms of other properties of the grammar. In order to find the motivation for the construction, Patten (2012a), for example, relates cleft constructions to the family of specificational

sentences and accounts for some of the IT-cleft's properties by looking at historical data and arguing in terms of inheritance.

With respect to the historical origin and rise of the IT-cleft construction various theories have been proposed, including influence from Celtic languages, where cleft constructions occurred earlier than in English, are more frequently used and occur with greater variation (Filppula 2009; Filppula/Klemola 2012). Another theory is based on the fact that the initial element lost its capacity of hosting contrastive constituents after the loss of V2 syntax. The cleft construction then evolved as a resolution strategy by manoeuvring "contrastive constituents in a position that fits the new, rigid SVO order while retaining their information-structural status" (Los/Komen 2012: 884). Another theory of the historical origins of the IT-cleft construction is provided by Patten (2012a). She argues that the IT-cleft acquired its construction-specific properties over time due to language change. According to Patten it inherits properties from various constructions, including specificational copular and predicate nominal constructions, definite noun phrases and restrictive relative clauses (2012a: 247).

As for pseudo-clefts, linguists disagree on whether there is also a predicational type in addition to the prototypical specificational type. A copular predicational sentence does not specify the value of a variable but predicates something of the subject noun phrase. An example of a predicational sentence is given in (3.69a). Compare the specifying pseudo-cleft in (3.69b).

(3.69) a. What they did was a disgrace.

b. What they did was paint their house red.

Some linguists argue that predicational sentences, such as (a), must be distinguished from pseudo-clefts (e.g. Ward et al. 2002), while others see predicational pseudo-clefts as just another type of pseudo-cleft (e.g. Declerck 1983, 1984). In the present study, only specificational pseudo-clefts will be discussed.

As for cross-varietal studies of cleft constructions, Mair and Winkle's (2012) ICE-based study deals with an ongoing change in a specific type of pseudo-cleft construction, namely the change from explicit marking to bare infinitive in sentences like *What they do is (to) travel around the world*. Comparing data from ten varieties of English, the study finds that Australian English, New Zealand English and Canadian English are most advanced in the development towards a preferred use of the bare infinitive, followed by British English and Irish English. Among the L2 English varieties analyzed, Philippine English is closest in its behaviour to the L1 English varieties and shows the highest use of bare infinitives, followed by Singapore English and Jamaican English. In Indian English and Hong Kong English, on the other hand, the infinitive marker *to* is still frequently retained.[43]

Lange (2012), in her analysis of the syntax of spoken Indian English, concentrates on IT-clefts. She notes that the construction occurs rarely in ICE-Great Britain and is an "even more marginal phenomenon" in ICE-India (2012: 178), where she finds a surprisingly low number of only seven instances of IT-clefts.

43 This specific type of pseudo-cleft is also mentioned in, for example, Allerton (1991), Collins (1991), Miller (1996, 2006), Miller and Weinert (1998), Rohdenburg (1998, 2000), Lambrecht (2001b), Calude (2009) and Rohdenburg and Schlüter (2009).

Siemund and Beal's (2011) study of IT-clefts is based on historical data (the Hamburg Corpus of Irish English and the Corpus of Late Modern English Prose) and present-day data (ICE-Great Britain, ICE-Ireland and ICE-India). The variation they find across varieties of English includes the following, among others. British English speakers use more interrogative clefts and they increasingly create more complex clefts (e.g. 'double topicalization': *Gary Stevens it was who came forward*; ICE-GB:S2A-001), innovations which, as Siemund and Beal argue, "suggest an ability and willingness to further complicate an already complex construction, which requires high proficiency and confidence levels in one's languages skills, as well as thorough understanding of the cleft construction both syntactically and semantically" (2011: 264). Furthermore, the study reveals differences between British, Irish and Indian English in terms of the syntactic function and the syntactic category of the clefted constituent. Finally, Irish English shows quite a number of subject clefts with zero subordinator, constructions which are considered to be completely intolerable by native speaker informants. Siemund and Beal explain that zero subordinators are heavily stigmatized features which have been used as markers of Irish nationality in the literature since the eighteenth century (2011: 255). Given that this type of IT-cleft is most frequent in the data of the two learner varieties Irish English and Indian English, Siemund and Beal argue that this is a case of speakers "striving for simplicity" (2011: 264).

Beal (2012) is an ICE-based study which compares IT-clefts in Irish English with those in British English, Jamaican English, Singapore English, Indian English and East African English. In contrast to the present study, Beal works with all files of the relevant ICE corpora and not only the direct conversation files. She finds that speakers of Irish English use IT-clefts as often as speakers of British English and consequently suggests that a "more precise analysis of Irish English ITCs [i.e. IT-clefts; CW] must be conducted in order to determine their supposed uniqueness beyond simple relative frequency" (Beal 2012: 161). She then examines IT-clefts according to genre and finds that the construction is generally used more frequently in speech than in writing; the only exception is ICE-Singapore where more items can be attested for the written data than for the spoken data. Furthermore, Beal analyzes the complexity and the syntactic function of the clefted element, and the animacy agreement between the clefted constituent and the subordinator, that is, "the prescriptive requirement for animate subordinators (*who*/*whom*) modifying animate clefted elements" (2012: 171). She finds that speakers of Irish English and Jamaican English show more flexibility with respect to the syntactic function of the clefted element than the speakers of the other varieties. Indian English has the highest proportion of subject clefts. Regarding animacy agreement, speakers of Irish English and Jamaican English deviate most frequently from prescriptive norms, while speakers of East African English produce by far the lowest variation from prescriptive norms.

3.4 Information packaging in some background languages

This section outlines the basic ways of structuring information in sentences in the (major) background languages of the varieties of English analyzed in the present study: Irish, Mandarin, Cantonese, Tagalog, Jamaican Creole, Hindi and Malayalam. In particular, possible counterparts to the English information-packaging constructions under consideration in the present study will be described. Many of the societies in which English is spoken as a L2 are, of course, multilingual. That is, many more languages than the ones discussed in this section are spoken in these communities. Reviewing all background languages would go beyond the scope of the present study. It was therefore decided to restrict the discussion of information-packaging strategies to the major background languages, those languages which can be assumed to have the strongest effect on the English variety spoken in these communities. In the case of India, Hindi and Malayalam are meant to represent the two major language families, namely Indo-Aryan languages and Dravidian languages.

3.4.1 Irish

Irish is the official language of the Republic of Ireland. Today, it is usually learned only as a second language at school, but when the English language came to Ireland, the population was monolingually Irish-speaking. Thus, the two languages have been in contact for several centuries and numerous traces of Irish can be found in Irish English.

Irish has a rigid word order VSO. Yet like in English, non-canonical order is found in constructions that are used for highlighting sentence constituents. The equivalent to the English cleft is a copular construction introduced by a form of copula *be*, followed by the element in focus and a relative clause. Contrary to the English cleft, however, there are hardly any constraints on clefting in Irish. With the exception of inflected verbs almost all constituents of the Irish sentence can be clefted. In addition to noun phrases, Irish freely allows prepositional phrases, adverbials and (uninflected) verb phrases to be clefted. The following examples, taken from Stenson (1981: 99), illustrate these different types. Note that Stenson marks some examples with question-marks or an asterisk, which means that the constructions are questionable or unacceptable in English.

(3.70) *is* *í* *mo* *dheirfiúr* *a* *chonaíonns* *i* *Sasana*
COP her my sister REL live-REL in England
'It's my sister that lives in England.'

(3.71) *is* *ar* *an* *mbóthar* *a* *bhuailfidh* *mé* *leat*
COP on the road REL meet-FUT I with-you
?'It's on the road that I'll meet you.'

(3.72) *is abhaile a chuaigh sé*
COP home REL go-PA he
?'It's home that he went.'

(3.73) *is ag déanamh a chuid ceachtannaí atá Tadhg*
COP at do-VN his portion lessons REL-be Tim
*'It's doing his lessons that Tim is.'

(Stenson 1981: 99)

As can be seen in example (3.70), clefted noun phrases must be preceded by a co-referential pronoun, agreeing in number and gender, if they are definite. That is, in the present case the definite noun phrase *mo dheirfiúr* 'my sister' requires the pronoun *í* 'her'. Such a co-referential pronoun is, in fact, required in all sentences where the copula precedes a definite noun phrase, not only in clefts. In the following sentence, *Tomás* 'Thomas' requires *é* 'him'.

(3.74) *is é Tomás mo dheartháir*
COP him Thomas my brother
'Thomas is my brother.'

(Stenson 1981: 96)

In addition to IT-clefts, Irish also has pseudo-cleft constructions, formed by the insertion of the noun phrases *an té* 'the one' or *an rud* 'the thing'. Note that in (3.77) the sequence *is é an rud* 'is it the thing' merges to *séard* 'what'.

(3.75) *is é an té a bhí tinn ná m'athair*
COP him the one REL be-PA sick namely my'father
'The one who was sick is my father.'

(Stenson 1981: 110)

(3.76) *is é an portach an rud a fheiceanns tú*
COP it the bog the thing REL see-REL you
'The bog is the thing that you see.'

(Stenson 1981: 113)

(3.77) *séard a fheiceanns tú (ná) an portach*
what COP see-REL you (namely) the bog
'What you see is the bog.'

(Stenson 1981: 111)

Stenson (1981: 111-113) argues that pseudo-cleft sentences such as those given above are structurally identical to identification sentences and that they underlie IT-cleft sentences. Thus, an IT-cleft sentence may be derived from (3.76), for example, by deletion of the noun phrase *an rud*. Hence, while IT-clefts are already emphatic constructions, pseudo-clefts with *séard* illustrated in (3.77) "are felt even more marked

('emphatic') than the cleft sentences they paraphrase" (Stenson 1981: 113). Stenson argues that this is because the initial sequence *is é an rud* fuses to *séard* and the highlighted element *an portach* 'the bog' is placed in final position, additionally set off from the rest of the sentence by the word *ná* 'namely'.

Word order variation, in particular fronting, leads to even more emphatic sentences. Fronting occurs in Irish in two types of copular sentences: in clefts with inversion and in classification sentences. Through the inversion of copula and predicate, that is, the initial placement of the predicate, the predicate is emphasized. These sentences are additionally marked by the presence of *ea*, "a remnant of an Old Irish neuter pronoun, now gone [...] from the modern language, except in this construction" (Stenson 1981: 116). Compare the following two sentences.

(3.78) *is* *pub* *maith* *é*
COP pub good it
'It's a good pub.'

(3.79) *pub* *maith* *is* *ea* *é*
pub good COP it it
'It's a *good pub*.'

(Stenson 1981: 116)

To illustrate the marking of emphasis by the cleft construction and its inverted version, Stenson gives the following three examples. While the cleft in (3.81) is considered more emphatic than the 'unmarked' sentence in (3.80), the inverted cleft in (3.82) shows "extra emphasis" as against the 'normal' cleft in (3.81).

(3.80) *tiocfaidh* *Dónall* *amárach*
come-FUT Donal tomorrow
'Donal will come tomorrow.'

(3.81) *is amárach a tiocfaidh Dónall*
'It's tomorrow that Donal will come.'

(3.82) *amárach is ea a tiocfaidh Dónall*
'It's *tomorrow* that Donal will come.'

(Stenson 1981: 117)

Turning to existential sentences, Irish has no equivalent to the English *there*-existential. To express existence Irish uses a construction consisting of a form of *be*, followed by a nominal, which in turn is followed by the element *ann*, a form of the preposition meaning *in* and translated as 'in it'. Consider the following examples for illustration.

(3.83) *tá* *fion* *ann*
is wine in-it
'There's wine.'

(McCloskey 2012: 36)

(3.84) *ní raibh aon ghluaisteán an-uair-sin ann*
NEG be-PAST any car that-time in-it
'There were no cars in those days.'

(McCloskey 2012: 6)

(3.85) *beidh go-leor bia ann*
be-fut plenty food in-it
'There'll be plenty of food.'

(McCloskey 2012: 6)

Irish forbids bare existentials or, more precisely speaking, Irish existentials always contain the predicative *ann*. That is, the sentence in (3.83) is a bare Irish existential; simply saying *tá fion* would be unacceptable.

Traces of Irish existentials can be found in Irish English existentials. The examples presented below show that the prepositional phrase *in it* is used to fulfil the functions of Irish *ann*. In (3.86a), the speaker was asked how many banks there were in Ballycastle. And in (b), the speaker talks about an old dress that does not fit her anymore.

(3.86) a. There's a Northern Bank in it (ICE-IRE:S1A-030)

b. There's no zip in it [unclear] I burst it the last time I put it on (ICE-IRE:S1A-063)

McCloskey (2012: 15) argues that the use of 'existential' *in it* is widespread in Irish English dialects. It has also been recorded by Filppula (1999: 228) and Bliss (1984: 149), among others.

3.4.2 Mandarin

Mandarin is the most widespread Chinese dialect in Singapore and has probably the greatest impact on the shape and development of Singapore English today (cf. Chapter 2, section 2.2.4). Mandarin differs typologically in a number of ways from English. The parameters that are most important in the context of the present study are word order and the basic orientation of the sentence (subject vs. topic). As noted earlier, Mandarin sentence structure is topic-prominent and thus differs from the subject-prominent sentence structure of English. Nearly all English sentences need a subject, which is usually easy to identify because it typically occurs before the verb and the verb agrees with it in number. In Mandarin, on the other hand, the syntactic category of subject is much less significant and it is the topic that plays a crucial role in the structuring of sentences. The subject of a Mandarin sentence is that noun phrase that has a 'doing' or 'being' relationship (like English subjects), yet it is not marked by a specific position or by agreement. The topic is what the sentence is about or, in Chafe's words, it "sets a spatial, temporal, or individual framework within which the

main predication holds" (1976: 50). Further properties of the topic are that it is definite or generic, it always occurs in sentence-initial position and it can be separated from the rest of the sentence by a pause or a pause filler, yet these topic markers are optional (Li/Thompson 1981: 85f.).

There are different sentence types that illustrate nicely the difference between the two notions in Mandarin. First, there are sentences with both topic and subject, as in (3.87).[44]

(3.87) *nèi zhī gǒu wǒ yijing kàn guo le*
that CL dog I already see EXP CRS
'That dog I have already seen.'

(Li/Thompson 1981: 88)

The topic *nèi zhī gǒu* 'that dog' occurs in initial position, is definite and specifies what the sentence is about. The noun phrase *wǒ* 'I' is the subject. An interesting type of sentence, which contains both topic and subject, is the double-subject construction (also cf. section 3.1). What is special about such sentences is that the topic and the subject stand in a part-whole relationship. Consider the sentence in (3.88) for illustration (example (3.5) in section 3.1).

(3.88) *nèike shù yèzi dà*
that tree leaves big
'That tree, the leaves are big.'

(Li/Thompson 1981: 94)

The topic *nèike shù* 'that tree' refers to the 'whole' and the subject *yèzi* 'the leaves' specifies a part of it.

As in English, topic and subject can also be identical in Mandarin. What is characteristic of Mandarin, however, is that there are many sentences that have a topic but no subject. In those sentences the subject is implicitly understood, as in (3.89).

(3.89) *nèi běn shū chūbǎn le*
that CL book publish PFV/CRS
'That book, (someone) has published it.'

(Li/Thompson 1981: 88)

Obviously, the book has not published itself or something else, but it is implicitly understood that some person or company has done so. Such Mandarin sentences are often translated as passive constructions in English. Note, however, that they are no passives in Mandarin but "simply topic-comment constructions in which the subject of the verb is not present" (Li/Thompson 1981: 89). The sentence in (3.89) illustrates a further interesting characteristic of Mandarin: pronouns that are co-referential with the topic are left out (cf. the pronoun *it* in the English translation). This applies

44 In example (3.87), *zhī* is a classifier (CL), the verb suffix *guo* indicates that the action has been experienced in the past (EXP) and the verb suffix *le* indicates that the action is completed (CRS).

even across sentence boundaries. Since the topic is what the sentence is about it is often the case that it persists in the discourse and that the ensuing sentences also say something about it. From this follows that the topic has priority over the subject in determining the co-reference of missing noun phrases in the following discourse. Consider the example in (3.90) for illustration.

(3.90) *nèi ke shù yèzi dà; (suóyǐ) wǒ bu xīhuān*
that CL tree leaf big (so) I not like
'That tree, the leaves are big; (so) I don't like it.'

(Li/Thompson 1981: 102)

This example illustrates that the topic can easily extend its scope across sentence boundaries while the subject's role is often restricted to the sentence in which it occurs.

A further distinctive property of Mandarin which follows from its topic prominence is that there is no equivalent to dummy subject *it*. In sentences where dummy *it* is required in English, it is simply left out in Mandarin. Compare the sentences in (3.91).

(3.91) *xià yú le*
descend rain CRS
'It's raining.'

(Li/Thompson 1981: 91)

Similarly, Mandarin has no equivalent to empty or existential *there*. Rather, it uses the verb *yǒu* or a "verb of posture", such as *zuò* 'sit', *tǎng* 'lie' or *piāo* 'float', to express the existence of something at a certain locus (Li/Thompson 1981: 510). As the following two examples illustrate, there are two ways of ordering the constituents in existential sentences: either the locus is the initial element or the verb *yǒu* comes first, followed by the entity whose existence is expressed. If the locus occurs in sentence-initial position, it "must be definite in the sense that its existence must have already been established in the discourse context either linguistically or extralinguistically" (Li/Thompson 1981: 511). And since the topic comes first in Mandarin sentences, the locus then takes on the function of the topic.

(3.92) *(zài) yuànzi li yǒu yi zhī gǒu*
at yard in exist one CL dog
'There's a dog in the yard.'

(3.93) *yǒu yi zhī gǒu zài yuànzi li*
exist one CL dog at yard in
'There is a dog in the yard.'

(Li/Thompson 1981: 510-511)

The verb *yǒu* can furthermore be used to express the existence of some entity in relation to some other entity. These sentences may express possession if the entity on which the existence of some other entity is predicated is animate. In the English translation the verb *have* is then used, as can be seen in the following example.

(3.94) *tā* *yǒu* *sān* *ge* *háizi*
3sg exist three CL child
'S/He has three children.'

(Li/Thompson 1981: 513)

Li and Thompson point out that the sentences with the verb *yǒu* given in (3.92) and (3.93) and that in (3.94) are essentially similar despite the difference between their English translations. In both types of sentence "something is being claimed to exist; the difference is whether it is said to exist with respect to a place or to another entity" (Li/Thompson 1981: 513).

In addition to *yǒu*, the copula verb *shì* can also be used to express existence, resulting in a difference in meaning, as the following contrasting pair shows.

(3.95) *wàimian* *yǒu* *yi* *zhī* *gǒu*
outside exist one CL dog
'There's a dog outside.'

(3.96) *wàimian* *shì* *yi* *zhī* *gǒu*
outside be one CL dog
'What's outside is a dog.'

(Li/Thompson 1981: 514)

The sentence in (3.96) implies that the locus is already known to the addressee and, additionally, that the speaker believes that the addressee is interested in the locus, what it has in it or what it looks like. The example with *yǒu* in (3.95), on the other hand, simply predicates the existence of some entity at some locus, here, the existence of a dog outside (Li/Thompson 1981: 515).

The example in (3.96) also shows that the copula *shì* can be used to form a cleft sentence, in this case a pseudo-cleft. The counterpart to the English IT-cleft is also formed with the help of *shì*, either with the copula alone or in combination with the particle *de*. Yet, note that the bare-*shì* sentence has different properties than the *shì…de* sentence. Consider the following examples.

(3.97) ***shì*** *Zhāngsān* *zuótiān* *kàndào* *Wáng* *xiǎojiě* *(bú* *shì* *Lǐsì)*
COP Zhangsan yesterday see Wang Ms not COP Lisi
'It is Zhangsan who saw Ms Wang yesterday (not Lisi).'

(3.98) ***shì*** *Zhāngsān* *zuótiān* *kàndào* *Wáng* *xiǎojiě* ***de***
COP Zhangsan yesterday see Wang Ms DE
'It is Zhangsan who saw Ms Wang yesterday.'

(Cheng 2008: 254-255)

In bare-*shì* sentences the constituent immediately following *shì* is the focused element and it always has a contrastive focus interpretation (cf. the addition of *bú shì Lǐsì* 'not Lisi' in (3.97)). On the other hand, *shì…de* sentences do not necessarily involve a contrastive reading (Cheng 2008: 255; Li 2008: 764).

The copula *shì* may also occur sentence-medially. Following the topic-prominent basic sentence structure, in (3.99) the topic *Zhangsan* comes first, followed by *shì*, which in turn is followed by the focused element.

(3.99) *Zhāngsān* ***shì*** *zuótiān* *lái* ***de***
Zhangsan COP yesterday come DE
'It was yesterday that Zhangsan came.'

(Hole 2011: 1707)

A more exact English translation of this sentence would be something like *As for Zhangsan, it was yesterday that he came.* [45]

3.4.3 Cantonese

Cantonese is a Chinese dialect, spoken as the first language by most people in Hong Kong. It is commonly assumed that all Chinese dialects share essentially the same grammar "apart from differences in suffix and particles for which, however, fairly close equivalents can be set up between dialects, one can say that there is practically one universal grammar" (Chao 1968: 13). Hence, everything that has been said on Mandarin Chinese sentence structure in the previous subsection basically also applies to Cantonese.[46]

Cantonese is an isolating language with the basic word order SVO. As in Mandarin, the basic sentence structure is topic-prominent. The topic does not need to be the subject of the sentence or bear any grammatical relation with the verb or the rest of the clause, as the follwing example illustrates.

(3.100) *gwo* *hói* *àh,* *deihtit* *jeui* *faai*
cross sea SFP underground most fast
'For crossing the harbour, the underground is fastest.'

(Matthews/Yip 1994: 78)

There is a semantic relationship between *gwo hói* 'crossing the harbour' and *deihtit jeui faai* 'the underground is fastest', which is a typical topic-comment relationship in Cantonese (Matthews/Yip 1994: 78). The ubiquity of such sentences makes Cantonese a topic-prominent language. From this property follow a number of grammatical features. For example, subjects must be definite, subject and object pronouns that refer back to a topic in the preceding discourse are often omitted ('topic chaining') and topicalization constructions are common (Matthews/Yip 1994: 78). Topicalization means that the topic is placed in initial position, usually for the purpose of contrasting the topicalized word or phrase with some other word. Topics can also occur in second position in the clause, as in (3.101).

---

45 This overview of sentences with *shì* and *de* somewhat simplifies the actual situation. The issue of *shì…de* clefts and *shì* clefts has received ample interest in the linguistic literature, with their structure and functions still not being agreed on. For more recent treatments see, for example, Cheng (2008), Paul and Whitman (2008), Li (2008) or Hole (2011).

46 Tang and Cheng (2014) discuss a number of structural differences between Cantonese and Mandarin, which are of no particular relevance for the present study though. These include structural particles, aspect markers, definiteness and double object sentences, among other things.

(3.101) *kéuih jyú faahn jeui lēk haih jīng yú*
s/he cook food most clever is steam fish
'As far as cooking is concerned she's best at steamed fish.'

(Matthews/Yip 1994: 88)

In this sentence, the secondary topic *jyú faahn* 'cook food' has less prominence than the initial topic *kéuih* 's/he'.

Corresponding to Mandarin *yǒu*, Cantonese has the verb *yáuh* to express possession and existence. Compare the following examples, whose structure is very similar to the Mandarin examples given above.

(3.102) *kéuihdeih yáuh sāam go jái*
they have three CL son
'They have three sons.'

(3.103) *lī go deihfōng yáuh mahntàih*
this CL place have problem
'There's something wrong with this place.'

(Matthews/Yip 1994: 317)

Cantonese has a focus-marking copula, yielding a construction equivalent to the English cleft. Corresponding to Mandarin *shì...de* sentences, Cantonese uses structures with *haih...ge* to focus on the entity after *haih* in the clause. Sentence (3.104) exemplifies a Cantonese IT-cleft.

(3.104) *gó bún syū haih ngóh dehng ge*
that CL book is I order SFP
'It was I who ordered that book.'

(Matthews/Yip 1994: 146)

In this sentence, the copula *haih* is used to emphasize the following word *ngóh* 'I'. The structure is also often used for contrastive purpose, much like the English cleft.

3.4.4 Tagalog

The national language of the Philippines, Filipino, is based on Tagalog, which belongs to the language family of Austronesian languages (along with Malay, Javanese, Sudanese, among many others). Tagalog is a VSO language and it is topic-prominent. This means that, in contrast to English, Tagalog basic sentences do not consist of subject plus predicate but of predicate plus topic (in this order; i.e. we have a comment-topic order). The topic expresses the focus of attention and in many cases corresponds to the subject of the English translation. Consider the examples in (3.105) and (3.106). In both cases the topics correspond to the subjects in the English sentences. The topics are identifiable through the topic markers *ang* and *si*, respectively.

(3.105) *kumuha* *ang* *bata* *ng* *kamatis*
got T-MKR child MKR tomato
'The child got a/some tomato.'

(3.106) *kumuha* *siya* *ng* *kamatis*
got T-MKR-he MKR tomato
'He got a/some tomato.'

(Rafael 1978: 36)

Such a correspondence does not always exist, however, and there are also many sentences where the Tagalog topic cannot be translated into an English subject or vice versa, that is, an English subject is not translatable into a Tagalog topic. For example, the English sentence 'A noise awakened the child' is in Tagalog *Ginising ng ingay ang bata*. In this sentence, the topic *bata* corresponds to the English object 'child' and the English subject 'noise' (*ingay*) is part of the predicate in the Tagalog sentence. The reason that the English subject 'noise' cannot be translated as the topic lies in the fact that it is indefinite and the Tagalog "topic never expresses a meaning of indefiniteness" (Schachter/Otanes 1972: 60).

In addition to the differences in position and definiteness, there is a third important difference between the Tagalog topic and the English subject. In English, the subject is typically associated with the semantic role of agent, but this is not the case with the Tagalog topic, which can occur in a variety of semantic relations to the verb, as the following examples illustrate.

(3.107) *sinulat* *ko* *ang* *liham*
write-PAST I the letter
'I wrote the letter.'

(3.108) *sinulatan* *ko* *ang* *titser*
write-to-PAST I the teacher
'I wrote to the teacher.'

(adapted from Schachter/Otanes 1972: 60)

The performer of the action is expressed by *ko* and is part of the predicate in both sentences (corresponding to the English subject 'I'). The topics *ang liham* and *ang titser*, on the other hand, have different semantic relations to the verb (cf. the differing forms of the verbs). Note, however, that they both express the focus of attention in the respective sentence.

Rafael (1978), furthermore, points out that different elements in the sentence may be chosen as the topic depending on the context and the sense and mood the speaker wants to convey. Compare the following examples, where topic choice determines the meaning of the sentence. There are two arguments, Jose and Maria, and Jose holds Maria by the hand. Rafael (1978: 41) argues that "[i]f Jose holds Maria to support, guide, or perhaps just to feel her, then Maria must be made topic", as in (3.109). By contrast, Jose is made the topic if he is supported or guided by Maria, as in (3.110).

(3.109) *hinawakan ni Jose si Maria*
held MKR Jose T-MKR Maria
'Jose held Maria.'

(3.110) *humawak si Jose kay Maria*
held T-MKR Jose MKR Maria
'Jose held Maria.'

(Rafael 1978: 41)

Similarly, in sentences with verbs such as 'admire', 'kiss', 'see', 'feel', 'hear', among others, the meanings conveyed are different depending on whether the experiencer or the patient is chosen as the topic (Rafael 1978: 42ff.).

In Tagalog, four different types of existential sentence can be distinguished. Minimally, they consist of the verb *may* 'exist' followed by a noun phrase, which in turn may be followed by a locative adverbial or a relative clause. This first type is illustrated in (3.111). In a second type, exemplified in (3.112), *may* is followed by the element *roon* 'there, in it'. Additionally, the notional subject is inflected with a linker (*-ng* or *na*). Instead of *may* or *mayroon* it is also possible to express existence with the element *magkaroon*, which is inflected for aspect, as shown in (3.113). The fourth type consists also of an inflected form of the element *magka*, but the element *roon* is absent.

(3.111) *may mga tao sa labas*
exist PL person LOC outside
'There are people outside.'

(3.112) *mayroon sa bahay na manok*
exist.there LOC house LK chicken
'There's chicken in the house.'

(3.113) *magkaka-roon ng isa-ng rebisyon ng libron-ng iyan*
ASP.exist.there NS one-LK revision NS book-LK this
'There's chicken in the house.'

(3.114) *nagka-(gera) sa Europe*
ASP.exist war LOC Europe
'There will be a war in Europe.'

(Sabbagh 2009: 678-679)[47]

As in many other languages, the verbs *may(roon)* and *magka(roon)* are also used to express possession.

(3.115) *mayroo-ng malaki-ng aso si Maria*
exist.there-LK big-LK dog S Maria
'Maria has a big dog.'

(Sabbagh 2009: 683)

47 The copular verb *gera* is in brackets because it is optional.

In order to highlight certain elements in a sentence, Tagalog uses a cleft construction (for argument focus) or a fronting construction (for adjunct focus). The following examples illustrate these two types of construction, respectively. The sentence in (3.116) is an answer to the question 'I heard that your motorcycle broke down', whereas (3.117) is an answer to the question 'When did your car break down?'. The focused elements are marked in italics.

(3.116) *hindi* *ang=kotse=ko* *ang=na-sira*
no ABS=car=my ABS=PF:broke down
'No. *My car* broke down.' (lit. 'No. What broke down is my car.')

(3.117) *kahapon* *na-sira* *(iyon)*
yesterday PF:broke-down (it.ABS)
'(It) broke down *yesterday*.'

(Nagaya 2007: 353)

These examples illustrate that prominence is achieved by placing the elements to be highlighted at the beginning of the clause (recall that in Tagalog basic sentences the verb comes first). It is also possible to place an element at the left periphery but outside the clause. This element normally establishes the topic of the sentence (Schachter/Otanes 1972: 485ff.). Such constructions correspond to the English left dislocation construction. Note that in Tagalog co-referential pronouns are optional, as the following example illustrates. The speaker is asked over the phone: 'Where are you now?'. The left-detached constituent is usually set off from the clause by a pause, indicated by the comma in (3.118).

(3.118) *kami,* *nasa* *labas=pa*
1PL.EXCL.ABS be.at outside=still
'As for us, (we) are still outside.'

(Nagaya 2007: 363)

The example also illustrates that the left-detached position is a preferred site for a contrastive topic (Nagaya 2007: 363). The topic is *we* and, in contrast to some other people, the group comprised by the pronoun *we* is still outside.

3.4.5 Jamaican Creole

Jamaican Creole, although no official language, is the de facto language of identity in Jamaica. Like English, Jamaican Creole is a SVO language but, as Christie (1996: 49) points out, it is "relatively topic prominent" simply because it has "not been much subject to the constraints of written language". Word order varies in cases of predicate clefting and other kinds of fronting. Jamaican Creole distinguishes between the clefting of predicative and non-predicative elements. Predicative clefting consists of

fronting an adjective or a verb, introducing it with the focus particle *a*. Consider sentence (3.119) for illustration.[48]

(3.119) *a* *sik* *Samwel* *sik*
HL sick S sick
'Samuel is really sick.'

(Bailey 1966: 86, quoted in Patrick 2007: 138)

Other elements that may be clefted include adverbials, nouns and pronouns. The focus particle *a* may again occur in clause-initial position.

(3.120) *a* *dat* *mi* *a* *tel* *yu*
HL that 1s PROG tell 2s
'That's what I'm telling you.'

(Patrick 2007: 140)

Existence is normally expressed with the invariant verb *(h)av* plus an indefinite pronoun subject, for example, *dem* 'them' or *yu* 'you'.

(3.121) *yu hav wan sinting niem Ruolin Kyaaf*
'There is something called Rolling Calf.'

(Patrick 2004: 422)

Furthermore, Jamaican Creole relative clauses show some noteworthy features. As in vernacular English, relative clauses may occur with zero object relative pronouns. Yet, zero subject relatives are also common (cf. (3.122)); they are in fact even preferred over overt marking in subject position (Christie 1996: 56).

(3.122) *so* *di* *wan* *woz* *gowin* *tu* *stiil* *it* *noo* *tek* *op* *wat* *him* *did*
do DEF one PAST going to steal it now take up REL 3s PAST
dig
dig
'So the one who was going to steal it now took up what he had dug.'

(Patrick 2007: 136)

Christie (1996), moreover, notes that relativization often co-occurs with fronting in Jamaican Creole, stating that "an extraordinarily high proportion of Jamaican relative clauses are constituents of NPs appearing in initial position in sentences which have an overall structure more appropriately describable as NP-NP-VP. In such cases the first NP represents a focused element" (1996: 48). Left dislocation is the focusing strategy she finds to be most closely associated with relativization.

(3.123) *entiting yu pudong dem waant it*
'Anything you put down, they want it.'

(Christie 1996: 52)

48 The focus particle *a* is also referred to as the 'highlighter'. That is why we find HL in the gloss in (3.119).

Christie gives a developmental account for the frequent co-occurrence of relativization and left dislocation. Relative clauses with resumptive pronouns, for example, represent an early stage in the evolution of Jamaican relatives, according to Christie (1996: 57-59).

3.4.6 Hindi

Hindi is the official language of India, along with English. In Hindi, an Indo-Aryan language, the basic word order is SOV, but "violations of normal order in the form of meaningful displacements of constituent [...] are an important syntactic feature" (Masica 1991: 394). The sentence-initial element is usually the topic, but it is not necessarily also the subject. Adverbials and even participle clauses can well be the topic in Hindi. Consider the following sentences for illustration. The particle *to* is sometimes used to mark the topic, as illustrated in (3.124).

(3.124) *aj* *(to)* *həm* *tenis* *zərūr* *kʰelẽge*
today (PTCL) we tennis certainly play.FUT.M.PL
'**Today** we will definitely play tennis.'

(Kachru 2006: 246)

(3.125) *əmrīka* *ja* *kər* *ucc* *sikşa* *prapt* *kərūga*
America go CP high education obtain do.1P.FUT.M.SG
'**Having gone to America** (I) will get higher education.' i.e.,
'I will go to America and get higher education.'

(Kachru 2006: 246)

The focus of the sentence normally falls on the element preceding the verb. Hence, if an element is moved into topic position it is usually de-emphasized. Consider example (3.126), where *sab log* 'everybody' is the focus (the element before the verb).

(3.126) *dhan* *sab.log* *cāhte* *haĩ*
wealth everybody want
'Wealth, **everybody** wants.'

(adapted from Masica 1991: 394)

An element under focus can also be identified by emphatic stress, focus particles or movement from its canonical position. The verb phrase, for example, can be emphasized by placing another element to the right of it, as in (3.127). If the verb phrase is placed in initial position some other element gains emphasis, as is the case with the term roṭī 'bread' in (3.128).

(3.127) *maĩne* *khāī-hai* *roṭī*
I eat-PAST bread
'I **ate** the bread.'

(3.128) *khāī-hai roṭī mai˜ne*
eat-PAST bread I
'I ate the **bread**.'

(adapted from Masica 1991: 395)

Unlike English, Hindi has no pleonastic subject construction to express existence. Rather, the difference between existential and non-existential sentences is realized by moving the subject noun phrase from its canonical position and by using demonstrative elements (Sinha/Thakur 2005: 250). Typically, the locative adverbial occurs in sentence-initial position in existential clauses, as illustrated in (3.129). By contrast, in the non-existential sentence the subject is in initial position.

(3.129) *jangal mẽ sher hai*
forest in lion be.PR
'There is a lion in the forest.'

(3.130) *sher jangal mẽ hai*
lion forest in be.PR
'The lion is in the forest.'

(Sinha/Thakur 2005: 250)

The following example, taken from Kachru (2006), also nicely illustrates the creation of existential meaning by the placement of the locative adverbial *mẽ* before the subject. Note that *raja* 'a king' is the focus of the sentence, immediately preceding the verb phrase.

(3.131) *kisī zəmane mẽ ek raja t^ha.*
some time.M.OBL in a king be.PAST.M.SG
'Once upon a time there was a king.'

(Kachru 2006: 252)

Note that both Masica (1991: 396) and Kachru (2006: 254) point out that the descriptions briefly outlined above represent only a preliminary account and that more work needs to be done on the constraints of displacement, and information structure in general, in Indo-Aryan languages.

3.4.7 Malayalam

Malayalam, along with Tamil, Telugu and Kannada, is one of the four major Dravidian languages. It is the statutory provincial language of the state of Kerala in southwest India, spoken as a mother tongue by 3.2% of India's population (2001 Census). The unmarked order of constituents in all sentence types is SOV yet there is considerable freedom of movement of constituents, which Asher and Kumari (1997: 1) find unsurprising because the function of a noun phrase is usually marked by a case marker or

postposition. For emphasizing constituents in a sentence, clefting is a very common strategy, which may also be accompanied by constituent movement (Asher/Kumari 1997: 181). Topics are primarily marked by their position at the beginning of the sentence. Because of the free movement of constituents other elements than the subject are allowed to occupy the topic slot. The following sentences are examples of a topicalized object (3.132) and an adverbial (3.133).

(3.132) *puuccaye* *ellaarum* *kuuti* *talli* *konnu*
cat-ACC all together beat-PP kill-PAST
'The cat, they all beat it to death.'

(3.133) *koottayatteekke* *naan* *kazinna* *aazca* *pooyirunnu*
Kottayam-ALL I last week go-PERF-PAST
'Kottayam, I went there last week.'

(Asher/Kumari 1997: 184)

Existence is expressed in Malayalam with the help of the copula verbs *aane* and *unte*. The two verbs have different meanings. The latter asks whether the entity is where it is expected to be and *aane* simply expresses the existence of the entity at some location. Compare the following sentences; also note the difference in meaning that is associated with different initial constituents.

(3.134) a. *aanakal kaattil aane*
'The elephants are in the forest' (and not anywhere else).

b. *kaattil aanakal aane*
'It is the elephants that are in the forest' (and not some other animals).

c. *aanakal kaattil unte*
'(The) elephants are in the forest' (where they are expected to be).

d. *kaattil aanakal unte*
'There are elephants in the forest.'

(Asher/Kumari 1997: 101-102)

CHAPTER 4

Analysis

This chapter presents the quantitative and qualitative findings of the present study. The five constructions under consideration - left dislocation, right dislocation, fronting, existential *there*-constructions and cleft constructions - will be dealt with separately in the following sections. This first paragraph is meant to give a first broad idea of the distribution of information-packaging constructions across the varieties of English analyzed. The driving questions of this first section are (1) in how far do the speakers of the different varieties of English analyzed use marked syntactic devices for structuring the information in a clause and (2) do they have certain preferences.

Table 4.1: The distribution of left and right dislocation, fronting, existential *there*-constructions and clefts in the S1A-files of nine ICE corpora (absolute token frequencies and normalized frequencies per 100,000 words).

corpus	LD		RD		fronting		existential		clefting	
	N	norm.	N	norm.	N	norm.	N	norm.	N	norm.
GB	72	35.7	90	44.6	35	17.4	682	338.2	428	212.3
IRE	129	64.0	124	61.5	51	25.3	746	369.9	450	223.1
NZ	95	41.3	83	36.1	36	15.7	600	260.9	448	194.8
CAN	97	46.0	35	16.6	28	13.3	573	272.0	471	223.6
SIN	117	57.6	48	23.6	53	26.1	417	205.1	351	172.7
PHI	169	78.0	30	13.8	25	11.5	578	266.8	490	226.2
JAM	169	79.2	19	8.9	24	11.3	481	225.5	531	249.0
IND	356	164.8	58	26.9	213	98.6	643	297.7	378	175.0
HK	170	71.4	26	10.9	14	5.9	591	248.3	194	81.5

Table 4.1 and Figure 4.1 show the frequencies of the five information-packaging constructions per 100,000 words[1] in nine ICE corpora, with Great Britain (GB), Ireland (IRE), New Zealand (NZ) and Canada (CAN) representing countries where English

1 The frequencies per 100,000 words were calculated by normalizing the absolute token frequencies attested for each corpus. That is, the absolute token frequencies were divided by the total number of words in the corpus and then multiplied by 100,000. The corpus sizes were computed with the help of the R package. For more information on the computation procedure see the section on the data and methodology, 1.2.

is spoken as a first language and the Philippines (PHI), Jamaica (JAM), India (IND) and Hong Kong (HK) representing countries where English is spoken as a second language. In Singapore (SIN), English is spoken at home by more and more people and it is learned as a first language by an increasing number of children. Hence, Singapore English can be classified as an L1 variety for more and more speakers.[2]

Figure 4.1 shows the frequencies per 100,000 words in a stacked form in order to give an overall impression of the five constructions which lead to a marked word order in English. Admittedly, this general overview somewhat simplifies the situation as there are, for example, different types of cleft construction to be distinguished. But still, I think this overview reveals some first interesting findings and raises a number of questions to be addressed in the following sections.

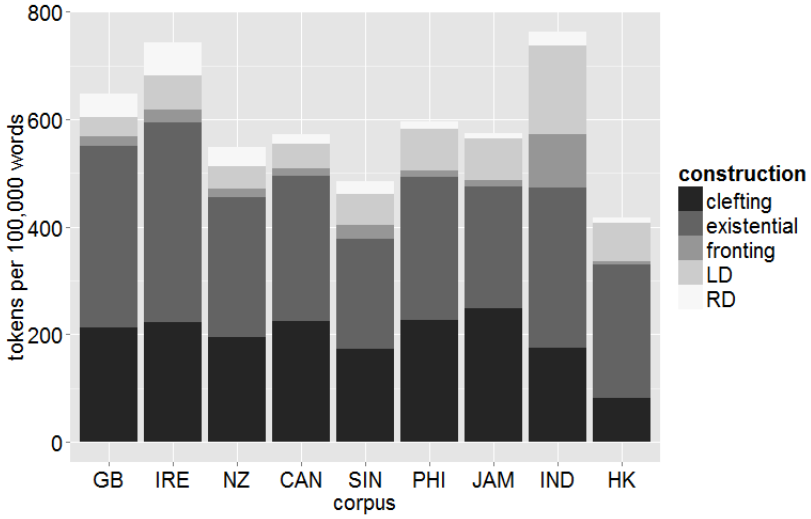

Figure 4.1: Information-packaging constructions in the direct conversation files of nine ICE corpora: frequencies per 100,000 words of left dislocation, right dislocation, fronting, existential clauses and cleft constructions.

First, it is interesting to note that Indian English speakers generally seem to use more marked constructions than the speakers of the other varieties, with the difference being made up in particular by the more frequent use of left dislocation and fronting constructions.[3] Second, and somewhat surprisingly, Irish English shows the second highest overall number of information-packaging constructions and thus stands out not only among the other L1 English varieties but also tops most of the L2 varieties. Third, dislocation constructions and fronting constructions are used much more rarely than existential clauses and cleft constructions. Fourth, while left dislocation seems to be a phenomenon rather found with L2 English speakers it is the other way round with right dislocation. This construction seems to be more common among L1

2 For practical reasons, the expressions 'component' and 'corpus' are used to refer to the 100 'private dialogue' files of the relevant ICE corpora only and not to the whole corpora (unless stated otherwise).
3 According to chi-squared tests, the difference is statistically significant at the level $p < 0.001$ for all varieties but British English (significant at $p < 0.01$) and Irish English (not significant).

English speakers than among L2 speakers. Fifth, the frequency of use of existential clauses is much more balanced than that of the other constructions analyzed. And sixth, cleft constructions are extremely rare among Hong Kong English speakers.

These general observations will be elaborated on in the following sections, addressing such questions as: (1) what motivates Indian English speakers to use fronting and left dislocation constructions so much more frequently than the other speakers do; (2) why do Hong Kong English speakers use cleft constructions so infrequently; (3) is the discrepancy between Irish English and British English - its input variety and very close neighbour - due to influence from Irish on Irish English; (4) may the overall lowest frequencies of information-packaging constructions in Singapore English and Hong Kong English be due to the fact that in both countries most people have a Chinese dialect as their L1.

4.1 Left dislocation

The general overview of information-packaging constructions in the previous section has already indicated that Indian English speakers use left dislocation constructions (LDs) more frequently than the speakers of the other eight varieties analyzed. This section will examine in some more detail the nature and usage patterns of LD tokens. In particular, it will try to find an answer to the question of what motivates Indian English speakers to use the construction so much more frequently. After a first general overview of the distribution of the construction, the following subsections will examine a number of syntactic and pragmatic properties. Furthermore, two specific types of LD tokens - *for*-LDs and pronominal LDs - will be investigated in the final subsections.

4.1.1 Overall distribution

Figure 4.2 shows that left dislocation constructions are overwhelmingly most common in Indian English (164.8 tokens per 100,000 words), with the frequency of use being more than twice as high as in most other varieties and even five times as high as in British English. The latter variety shows the lowest frequency of use (35.7 per 100,000 words), followed by New Zealand English (41.3) and Canadian English (46.0). Irish English, on the other hand, sticks out among the L1 English varieties and, interestingly, shows an even higher frequency than Singapore English (64.0 vs. 57.6). Chi-squared tests were performed to test whether the probabilities of LD tokens in Indian English and the other varieties are essentially the same (as estimated by their frequencies). This is not at all the case as the highly significant results for all varieties show ($p < 0.001$).

Given these observations, two interesting points are worth mentioning. First, the

data suggest that it is not only fronting constructions and IT-cleft constructions that are commonly used by Irish English speakers, as has repeatedly been stated in the literature (cf. e.g. Filppula 1999, 2009, 2012; Hickey 2005, 2007), but that left dislocation can also be seen as a distinctive feature of this English variety, at least in comparison to the other L1 English varieties.[4] Second, assuming that LD is a learner feature, the relatively low frequency of use in Singapore English and the resulting similarity to the L1 varieties British English, New Zealand English and Canadian English may be a sign of the high proficiency of its speakers and the fact that more and more English speakers in Singapore acquire the language as their L1. Furthermore, this observation substantiates the claim that Singapore English can be seen as a stable nativized variety of English (Phase 4 of endonormative stabilization in Schneider's 2007 Dynamic Model). Among the L2 English varieties analyzed in the present study, Singapore English has been categorized as being most advanced towards endonormative stabilization. Note, however, that the differences between Singapore English and the L2 varieties are not statistically significant (except for Indian English, as shown above).

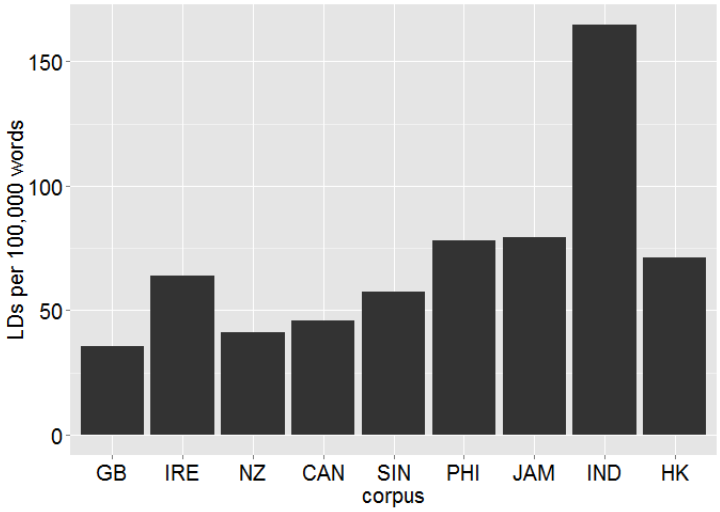

Figure 4.2: Left dislocation in nine ICE corpora (frequencies per 100,000 words).

With the exception of Singapore English and Irish English, Figure 4.2 also indicates that LD constructions are generally more common in L2 English varieties than in L1 varieties. This result supports the claim that LD tokens are frequently found in the speech of learners of English, a claim that has been put forward in various fields of language research, for example in language teaching and first and second language acquisition (cf. e.g. Gruber 1967; Chambers 1973; Cotton 1978; Williams 1987; Mesthrie 1992; Carter/McCarthy 1995; Ortega 2009). It has been suggested that the construction marks a transitional stage on the way to full knowledge of the language, which makes sense as the breaking down of a clause into smaller chunks eases the production and processing of the utterance. LD constructions are therefore pre-

4 According to chi-squared tests, the differences in frequency are statistically significant for British English ($p < 0.01$) and New Zealand English ($p < 0.05$) but not for Canadian English.

ferred among learners.

However, besides the shared L2 effect there must be other forces at work as well, otherwise the overwhelmingly more frequent use in Indian English and the relatively higher number of occurrences in Irish English (as compared to the other L1 varieties and Singapore English) would remain unexplained. The following paragraphs will discuss possible explanations for the high frequencies of LD tokens in these two varieties of English. Additionally, a number of idiosyncratic and noteworthy features relating to syntax and discourse will be analyzed.

4.1.2 Topic continuity

Examining Indian English data, Lange (2012: 165ff.) suggests that one explanation for the high frequency of LD tokens might be the specific discourse function of creating topic continuity, which has so far not been attested for other varieties of English, according to Lange. In LD constructions that serve this function, the salient noun phrase of the immediately preceding utterance is repeated in the dislocated element. Consider the following examples, which illustrate the typical discourse context in which LD tokens with this specific function occur. In (4.1), three people talk about different types of oil and discuss the question of which oil is best for fish preparation. Note that there are two examples of LD tokens in which the salient noun phrase of the immediately preceding utterance is taken up in the LD construction, *refined oil* and *taste*.

(4.1) D: If you want to make a fried fish <,> what you do

A: You just fry it <,>

D: Fry it in mustard oil

A: Mustard oil

We use mustard oil for <,> fish preparation I mean even the newer generation which is using uh refined oil <,> uhm it uses mustard oil for the <,> fish preparation

D: You can use **refined oil** for whatever <,>

A: Yes you know **refined** you know postman or saffola whatever you use it's of no use <,>

No

D: It won't allow the taste to come through

C: It won't allow **the taste** to come through

A: **The taste** it is very flavour but they say you've lot of cholesterol there <,>

(ICE-IND:S1A-007)

In example (4.2), two women talk about the Indian cricket team, which has lost most of its past matches. They agree that it must be the lack of team spirit which is responsible for their unsuccessful playing.

(4.2) A: Something is wrong with the <,> yeah <,> Indian team <,>
I think they don't have that uh <,> **team spirit** or something
B: **Team spirit** that's lacking <,,>
A: Because uh
B: I mean and that's why they keep losing

(ICE-IND:S1A-014)

In the following example, three students talk about how they spent their summer holidays.

(4.3) B: Tell me did you go anywhere in this **summer** <,> ?
C: How did you spent your **summer** time?
A: **Summer** I don't know how it passed <,,>
I just went to Madras for a short visit <,,>

(ICE-IND:S1A-031)

In LD constructions it is often the case that the theme of the previous discourse is taken up by the initial noun phrase or that the initial element is discourse-old. Yet, what is special about the examples given in (4.1) to (4.3) is that the salient noun phrase of the IMMEDIATELY preceding utterance is taken up and placed in initial position. The speakers uttering such sentences signal that they have carefully listened to the interlocutors, they are being polite and show that they have understood what the others have just said. Such LD tokens are thus used to create coherence in discourse, organize the information flow and also to empathize with the interlocutor. They are often found after questions, as in (4.3). In these cases, the speakers not only signal that they have understood the question, but they also gain time to find an answer.

The findings of the present study cannot fully confirm Lange's (2012) claim. In ICE-India, there is indeed a substantial number of LD tokens that serve the function of creating topic continuity (9.8% out of all LD tokens). However, this function is not exclusive to Indian English, as can be seen in Table 4.2. The rows called 'TC tokens' and 'TC %' (TC = topic continuity) present the absolute numbers and proportions of LD tokens that serve this specific function, respectively. It can be observed that the proportion of such LD tokens is similarly high in ICE-Ireland (9.3%) and even higher in ICE-Philippines (10.6%).

Table 4.2: LD tokens that create 'identity links' (absolute frequencies and percentages out of all LD tokens).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
LD tokens	72	129	95	97	117	170	169	356	170
TC tokens*	4	12	6	1	7	18	9	35	8
TC %	5.6	9.3	6.3	1.0	6.0	10.6	5.3	9.8	4.7

*TC = topic continuity

Given these numbers, it is not satisfying to refer only to the function of creating topic continuity as the major motivating factor for Indian English speakers to use LD constructions more frequently than the speakers of the other varieties; at least if the notion of 'creating topic continuity' is seen in this narrow sense referring to this specific type of LD construction only. What is worth doing, however, is giving the notion a wider interpretation and classifying all LD tokens whose first element refers to known or old information as 'creating topic continuity'.[5] The discussion in the following paragraphs will concentrate on this idea.

In order to get a better idea of the LD construction's discourse functions and to test the claim that Indian English speakers use the construction not so much for the introduction of new topics but rather to pick up old topics and thus create topic continuity, all items were classified according to the anaphoricity and the topic persistence of the initial element.[6] The investigation of the anaphoricity ratings is also interesting in so far as there is no agreement in the literature about the information status of the dislocated element. Many researchers claim that the referent of the initial element is usually already known or inferable from the context (e.g. Givón 1976, 1993; Lambrecht 1994), while others state that LDs may also serve to introduce new referents (e.g. Prince 1998; Birner/Ward 1998).

Both features - anaphoricity ratings and topic persistence scores - comprise three classes. As far as anaphoricity is concerned, we have LD tokens whose initial items refer to new information, that is, items that have not previously been mentioned in the discourse; second, there are items referring to old or known information, that is, items which have been mentioned in the previous discourse; and finally there are items which have not explicitly been mentioned but are inferable from the discourse because they are members of a previously mentioned set.[7]

Topic persistence refers to the question whether and in how far the initial element of a LD construction persists in the following discourse. In order to rate the persistence of the initial elements of the LD tokens, the following five sentences were looked at and the items grouped into LD tokens whose initial element is not mentioned in the following five sentences at all or is mentioned again in the following turns in form of a pronoun. It is assumed that pronominally taken up constituents are well established topics. The third group contains all items that do not belong to any of the other two groups, that is, items that are taken up as a noun phrase in the following discourse or items that are only indirectly talked about afterwards. This means that there is some indirect connection between the theme of the following discourse and the preclausal

5 Note that the terms 'old', 'given' and 'known' are here used interchangeably. For meaningful distinctions between these terms see, for example, Erteschik-Shir (2007), who distinguishes between 'old' and 'given' information, and Prince (1978), who draws a distinction between 'given' and 'known'.

6 The idea of the anaphoricity ratings and topic persistence scores is based on Gregory and Michaelis (2001).

7 Givenness is here understood in the sense of discourse-old/new information. This seemed to be the option that can best be operationalized in the present context. Recall that the linguistic term 'givenness' has extensively been discussed in the literature and much more fine-grained definitions have been suggested (cf. section 3.1). Yet, relying on speakers' assumptions or addressees' knowledge makes the givenness notion impossible to work with in the present context since I obviously have no direct access to the speakers and addressees' knowledge; I can only judge from their utterances and the context.

constituent. Consider the following example for illustration. The LD construction is formed around the entity *your parents*, but speaker A then continues to ask a question about the father. There is obviously a connection between the preclausal element *your parents* and the theme of the following discourse, *your father*.

(4.4) A: Now **your parents** are they alive and well
B: Uh yes
A: How How old is **your father**
B: Uhm <,> I don't know exactly
He 's coming up to sixty <,>

(ICE-GB:S1A-051)

Figure 4.3 plots the anaphoricity ratings of LD tokens in the varieties of English analyzed.[8] As explained above, the three categories contain items that are recoverable through explicit mentioning in the preceding discourse ('old') or through inferential linking ('inferable'), or they are not recoverable from the previous discourse ('new'). The relevant proportions are given as percentages out of all LD tokens in the respective variety of English.

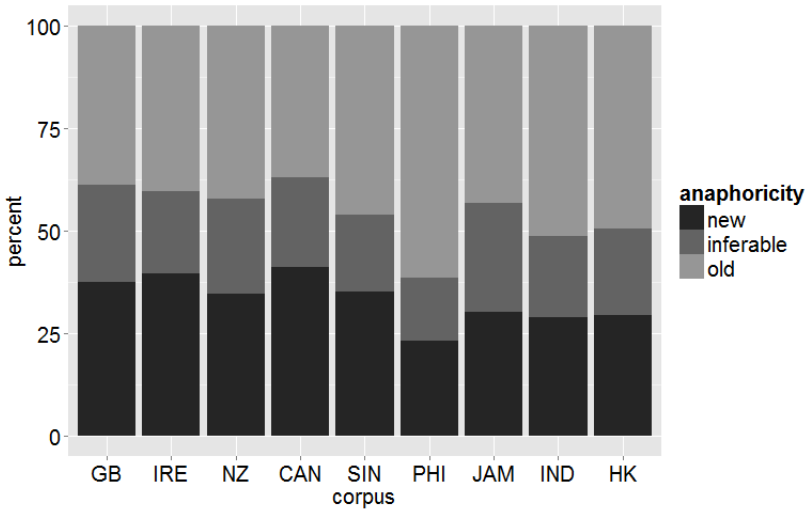

Figure 4.3: Anaphoricity ratings of LD tokens (percentages of initial elements containing old, new or inferable information).

The findings suggest that there is a L1-L2 divide as far as the information status of the initial elements is concerned. As can be seen, while initial elements containing inferable information make up the smallest proportion in all varieties analyzed, there is a slightly higher proportion of items containing new information in the L1 English varieties and Singapore English than in the other L2 varieties. In the latter, on the other hand, we find a high proportion of initial elements containing old information.

8 The token frequencies and percentages corresponding to Figure 4.3 are given in Appendix 6.6.5.

The proportion of such items is especially high in Philippine English, accounting for more than 61% of all LD tokens.

Another important observation to be made is that Indian English speakers do not behave particularly differently from the speakers of the other varieties. That is, the proportion of preclausal elements representing known information is only slightly higher than in most other varieties. Rather, it is Philippine English speakers who use LD tokens referring to old information very frequently. For them, the major discourse function of LD tokens seems not to be the introduction of new entities into the discourse but rather to put emphasis on certain elements in the clause or to contrast them with some other entity in the discourse. What further contributes to the high proportion of LD tokens whose initial element refers to known information is a specific type of LD token that is particularly common in Philippine English. In this specific type of construction the initial element has the form of a pronoun, which obviously refers to known information (*You know I'm open to anything you know **me** I'm a team player*. ICE-PHI:S1A-026). This specific type of LD token will be discussed in some more detail below. Note that ICE-Philippines still shows the highest proportion of LD tokens whose initial element refers to old information even if these pronoun LDs are excluded from the counts.

As for the information status of the initial element, it has been noted above that researchers take different positions in this respect. While some claim that the item usually contains known or inferable information, others also allow items referring to new information to be left-dislocated. The findings of the present study indicate that LD tokens serve both functions. Typically, the initial elements are known or inferable (cf. categories 'old' and 'inferable') in all nine varieties of English analyzed, yet the proportion is somewhat higher in the L2 varieties. Singapore English is closer in its behaviour to the two L1 varieties. Hence, we might speak of a L1-L2 divide with Singapore English on its way to becoming a L1.

Coming back to Indian English and the hypothesis that the function of creating topic continuity is a major motivating factor for its speakers to use LD constructions more frequently than the speakers of the other varieties do, this hypothesis cannot be supported by the analysis of the information status of the initial elements. Although Indian English speakers use LD tokens referring to old information more frequently than most other speakers (except for Philippine English speakers), really supportive evidence in favour of the hypothesis should yield a more striking difference. Yet, what also needs to be considered when talking about topic continuity is the ensuing discourse. This will be done in what follows.

Examining the topic persistence of the preclausal element in LD constructions, it can be said that a topic is well introduced and established in the discourse if it is pronominally taken up in the following turns. Assuming that LD constructions are often used to establish a topic or to mark a topic as salient, we would expect to find high scores of pronominal references in the following utterances or that the topic is at least indirectly continued. This expectation is met by all varieties of English analyzed, as the high proportions of the categories 'pronoun' and 'NP/indirect' in Figure 4.4 indicate.[9] Taking these two categories together, the highest proportions of continued

9 The results for Hong Kong English need to be taken with a pinch of salt because the Hong Kong ICE component contains the speech of many speakers who do not qualify as speakers of Hong Kong

topics can be attested for Irish English (81.4%) and New Zealand English (81.1%).[10]

As for Philippine English, it shows the highest proportion of pronominal referents in the ensuing discourse (58.6%). A contributing factor to this high proportion is again the specific type of pronoun LD mentioned before (*me I* ...) because in these cases the initial pronoun is often also mentioned in the following five sentences. Pronominally referred to entities are least frequent in the speech of Indian English speakers, accounting for less than half of all LD tokens (43.5%). The variety shows a somewhat higher proportion of items of the category 'NP/indirect', that is, LD tokens whose initial element is either taken up as a noun phrase in the following discourse or is indirectly referred to as part of a broader discourse topic.

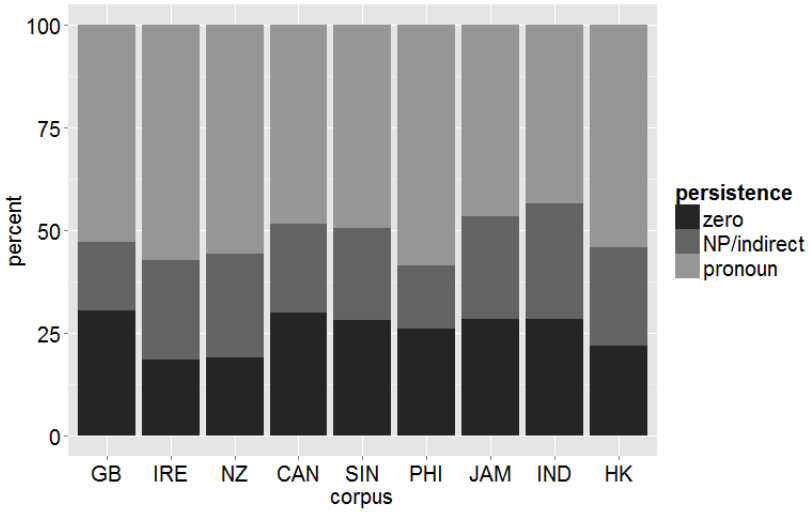

Figure 4.4: Topic persistence scores of LD tokens (percentages out of all LD tokens).

Given these results, we must conclude that the analysis of the persistence of the initial element in the following discourse does not yield evidence in support of the hypothesis concerning Indian English speakers and the motivating factor of creating topic continuity. Supportive evidence would have involved higher proportions of items that persist in the form of a pronoun in the following discourse. Another possibility that could still support the hypothesis would be a higher proportion of topics that are continued in the form of noun phrases (these items are included in the category 'NP/indirect'). However, this turned out not to be the case, that is, Indian English speakers do not use continued noun phrases more frequently than the speakers of the other varieties.

Summing up, this section set out to find evidence in support of the hypothesis that

English. Consequently, their contributions to the conversations have to be excluded from the analysis of Hong Kong English. This has been done in all the other investigations but is not possible in the present context. Excluding the turns of non-Hong Kong speakers and jumping to the next Hong Kong speaker turn would interrupt the information flow. Thus, it was decided to look at the following five sentence after an LD token regardless of who uttered these sentences.

10 The token frequencies and percentages corresponding to Figure 4.4 are given in Appendix 6.6.5.

the discourse function of creating topic continuity is a major motivating factor for Indian English speakers to use LD tokens. It was shown that the mere investigation of LD tokens which repeat the immediately preceding topic noun phrase in initial position, as suggested by Lange (2012), is not enough. If we compare only British English and Indian English, this may lead to convincing results, but if more varieties of English are taken into consideration, the picture gets more complex and the results less straightforward. The investigation of LD tokens that create 'identity links' yielded no significant results that could convincingly account for the high frequency of LD tokens in ICE-India. The additional examination of the anaphoricity and persistence of the preclausal element was to shed more light on the LD tokens' discourse functions and the motivations of speakers to use this marked way of organizing the information in a clause. Yet, as has been seen, neither the anaphoricity ratings nor the topic persistence scores yielded supportive results for the hypothesis. Supportive results for the hypothesis would have involved higher scores of initial elements containing old information and higher scores of pronominally or nominally referred to referents in the following discourse, but this is not at all the case. We do indeed have a somewhat higher proportion of LD tokens that refer to old information in Indian English, but the difference to the other varieties is so small that the finding is not really supportive of the hypothesis. For the topic persistence scores even the opposite is the case: Indian English shows the smallest proportion of pronominally referred to referents. Preclausal elements that are continued as noun phrases in the following discourse also did not turn out to be particularly common in the speech of Indian English speakers in comparison to the speakers of the other varieties.

Yet, we have seen that other varieties of English show a somewhat deviant behaviour, for example, Philippine English with its high number of pronoun LDs. More on this below.[11]

4.1.3 Simplifying function

As noted earlier, LD constructions also serve a simplifying function in the sense that they facilitate the online production and processing of an utterance by breaking it down into smaller chunks (cf. section 3.2.1).[12] If this is an important function of the construction, one could expect to find more LD tokens with complex initial elements rather than simple ones. In order to test this hypothesis the complexity of the preclausal elements was examined and compared across varieties of English. In the present context, the notion 'complex' refers to preclausal elements that either have the form of a clause (finite or non-finite) or a noun phrase postmodified by a relative

11 Anaphoricity ratings and topic persistence scores have also been calculated for fronting constructions. This allows for a direct comparison of the discourse functions of the two types of construction. While they are structurally quite similar, the anaphoricity ratings and topic persistence scores show that they perform different functions in discourse. A closer inspection of the differences is given in section 4.3.

12 Note that in this section the 'simplifying' function does not refer to Prince's (1998) notion but is understood in the sense employed by Huddleston/Pullum (2002) and Biber et al. (1999), namely that of making the planning and processing of complex sentences easier. See section 3.2.1 for more information about these different interpretations.

clause. Examples of these types of construction are given in (4.5).

(4.5) a. And she knows that usually **after we eat** that's the time we go (ICE-CAN:S1A-028)

b. Because **to make a pizza** it costs you like ten p and you sell it for fucking two quid (ICE-IRE:S1A-010)

c. But but **working <,> in this group** uhm <,,> it's <,> different in terms of uhm <,> the way <,> that you have to dance (ICE-GB:S1A-002)

d. **Most pagans who get married** yeah they have a pagan ceremony (ICE-GB:S1A-071)

Figure 4.5 indicates that there is quite some variation in the frequency of complex LDs across varieties of English.[13] As can be seen, the L1 varieties show much higher frequencies of complex LD tokens than the L2 varieties, with the exception of Jamaican English, which - quite surprisingly - has the third highest proportion of complex tokens (27.2% out of all LD tokens). The highest proportion can be attested for British English (31.9%), followed by Canadian English (28.9%).

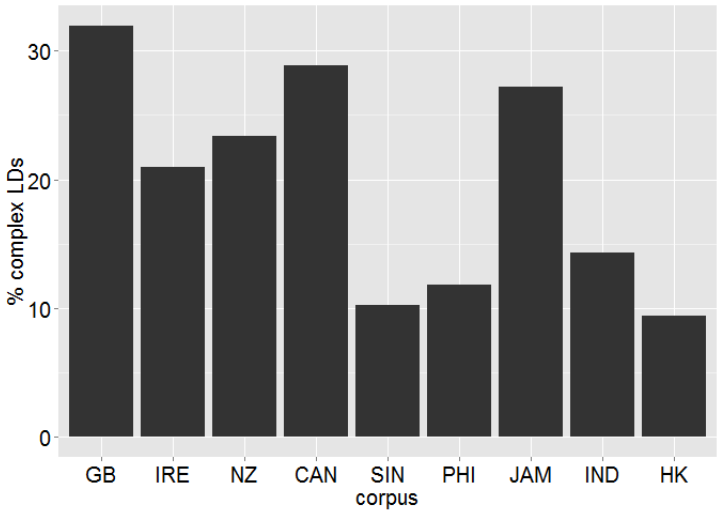

Figure 4.5: Proportion of LD tokens with complex initial elements.

Given these numbers two aspects are worth mentioning in particular. For one, Jamaican English sticks out among the L2 varieties and the high proportion of complex LDs is striking. A possible explanation for this high proportion might be that Jamaican English speakers generally tend to use relative clauses more frequently than the speakers of the other varieties. Evidence in support of this claim derives from the investigation of complex existential clauses, that is, existential *there*-constructions in which the notional subject is postmodified by a relative clause or a nonfinite clause.

13 The token frequencies and percentages corresponding to Figure 4.5 can be found in Appendix 6.6.5.

Compared with the other varieties of English analyzed, this type of construction is also very common in Jamaican English (for more details see section 4.4).

The second aspect worth discussing is the high proportion of complex LDs in the L1 varieties in comparison to the L2 varieties (except for Jamaican English). This finding may suggest that L1 English speakers are more likely than L2 English speakers to resort to LD constructions as a simplifying strategy, that is, the simplifying function is a more important discourse function for L1 English speakers while other functions are more important to L2 English speakers. This would be in line with the previous investigation of the anaphoricity of the preclausal element. There, the findings suggested that the creation of topic continuity and the organization of the information flow seemed to be a more important discourse function for the L2 speakers than the L1 speakers (recall that L2 speakers tend to use LD constructions referring to known information more frequently than do the L1 speakers).

In addition to the LD construction's discourse functions, the present study also investigated a number of syntactic features (still) aiming at finding possible explanations for the high frequency of LD tokens in Indian English. Furthermore, this was to reveal possible idiosyncratic features and similarities/differences in terms of preferred variants across the varieties of English analyzed. First, the syntactic function of the preclausal element or its co-referential pronoun is examined. This is followed by the investigation of LD tokens containing a demonstrative pronoun or a quantifying expression in the initial element.

4.1.4 Syntactic features

Syntactic function of the preclausal constituent

Examining the syntactic function of the initial element, we typically find items that are co-referential to subject pronouns in the core of the clause (cf. (4.6)a), but there are also quite a number of co-referential object and possessive pronouns. Adverbials, complements and verbs occur very rarely in dislocated position. Examples of these different types of LD construction are given in (4.6).

(4.6) a. **My ex-boyfriend Phil** [he] got me interested (ICE-GB:S1A-081)

b. **My Mum's Mum** I love [her] to death (ICE-IRE:S1A-005)

c. **Tommy Brick** [his] Granda died (ICE-IRE:S1A-005)

d. **The last meeting** I wasn't [there] (ICE-SIN:S1A-045)

e. **Athlete and sportsman** [that] he is (ICE-NZ:S1A-090)

f. **Show off** they'll do (ICE-IND:S1A-053)

The preponderance of subject LDs in all corpora analyzed is not particularly surprising since objects can alternatively be placed in initial position by means of fronting constructions. Simply moving subjects at the beginning of the clause in this way obviously has no specific pragmatic effect as subjects are typically found in initial

position in English.

It could be hypothesized that maybe speakers of Indian English are more flexible with respect to the syntactic function of the dislocated element and allow for a wider scope of realizational options than the speakers of the other varieties analyzed. If they put in initial position elements other than subjects more frequently, this in turn could account for the high frequency of LD tokens in Indian English. However, as can be seen in Table 4.3, this is not at all the case. It is rather the other way round and Indian English shows the highest proportion of subject LDs among all nine varieties of English analyzed, accounting for more than 89% of all LD tokens.

Table 4.3: The distribution of LD tokens according to the syntactic function of the initial element (absolute token frequencies and percentages out of all LD tokens.

corpus	LD tokens	subject		object		poss.		other*	
	N	N	%	N	%	N	%	N	%
GB	72	59	81.9	10	13.9	3	4.2	-	-
IRE	129	92	71.3	23	17.8	11	8.5	3	2.3
NZ	95	70	73.7	18	19.0	5	5.3	2	2.1
CAN	97	77	79.4	13	13.4	7	7.2	-	-
SIN	117	101	86.3	13	11.1	2	1.7	1	0.9
PHI	169	138	81.7	26	15.4	5	3.0	-	-
JAM	169	141	83.4	20	11.8	8	4.7	-	-
IND	356	318	89.3	34	9.6	2	0.6	2	0.6
HK	170	149	87.6	17	10.0	1	0.6	3	1.8

*The category 'other' comprises adverbials, complements and verbs.

The four L1 English varieties show more flexibility than the L2 varieties and form LD constructions around non-subjects more frequently, especially so Irish English and New Zealand English. Interestingly, it is exactly these two varieties of English that show the lowest proportion of fronted objects by means of a fronting construction which will be discussed in section 4.3. With fronting constructions we typically find fronted objects rather than complements or (obligatory) adverbials. It seems that speakers of Irish English and New Zealand English deviate more freely from the default mappings in both cases (subject LD and object fronting) than the speakers of the other varieties, followed by the two other L1 varieties, Canadian English and British English.

The findings suggest that there is a correlation between the syntactic function and the LD construction's major discourse function, namely topic establishment. As has already been pointed out in the chapter on information packaging (cf. Chapter 3), the subject of an unmarked declarative sentence in English is typically interpreted as the topic. Subjects are the prime candidates for the position of topic because they usually encode easily accessible language data. This is due to the fact that subjects normally have human referents and most often play the semantic role of agent (Seoane 2006: 366ff.). These three features - animacy, agentivity and high degree of accessibility - in turn, make subjects easy to process on the part of the hearer. Hence, it makes sense

that most LD constructions are formed around a subject in all nine ICE corpora: subjects are the prime candidates for topichood and LD tokens mainly serve the function of topic establishment or the marking of a topic as salient in the discourse. The especially high proportion of subject LDs in the L2 English varieties, and in particular in Indian English, might be seen as further evidence of the claim that the creation of topic continuity is a more important motivation for L2 English speakers to use LD constructions than for L1 English speakers. Also, it shows that the simplifying function plays no major role for L2 speakers (also see the relatively low proportion of complex LDs discussed above).

As for the somewhat higher proportion of possessive LDs in Irish English as compared to other the varieties, I can only offer a tentative explanation. The somewhat higher incidence of such items may be due to influence from the Irish substrate because resumptive possessive pronouns are also found in Irish passive and progressive constructions. Consider the examples in (4.7) and (4.8).

(4.7) *tá mé do mo bhualadh*
is I to my beating
'I am being beaten.'

(4.8) *tá sé ina shuí*
is he in.his sitting
'he is sitting.'

(Pietsch 2008: 216-217)

The structure of these sentences resembles possessive LDs to a great extent. It is likely that the Irish passive and progressive constructions raise the awareness of the existence of resumptive possessive pronouns and make them more accessible to speakers of Irish English. Consequently, Irish English speakers may be more prone to use them in other constructions as well.

LDs with a quantifying expression

The following paragraph will deal with another formal feature of the initial element in LD constructions for which some distributional differences can be attested across the varieties analyzed. The LD tokens in question involve a quantifier in the preclausal element, as illustrated in the sentences in (4.9).

(4.9) a. **All these northerners westerners** <,> they prefer wheat (ICE-IND:S1A-072)

b. And not only that but also **a lot of this stuff** that is to be done I think it requires some sort of uh <,> manpower and you'd more get that from a guy [laughter] than a woman (ICE-JAM:S1A-007)

c. But actually you know uh **many uh foreigners** they like this kind of uh Chinese uh porcelain china (ICE-HK:S1A-009)

d. I have some friends come from China they told us that uh you know **most of the people** you know they really don't believe in communism [...] (ICE-HK:S1A-009)

e. But **some uh people** they already got their privileges (ICE-HK:S1A-009)

These examples illustrate the most common types of quantifiers that occur in such a way in LD constructions, namely *all*, *a lot*, *many*, *most* and *some*. It could, of course, well be that the variation in LD tokens containing a quantifying expression is due to an overall higher frequency of use of these quantifying expression in one or the other variety of English analyzed. Hence, to get a better idea of the overall distribution of these quantifying expressions all items were counted in the nine corpora. Table 4.4 gives the absolute token frequencies and the normalized frequencies per 100,000 words of these five quantifying expressions (cf. the column labelled 'quantifier') together with the absolute frequencies of LD tokens that have an initial element with a quantifier (cf. 'quantifying LDs N'), the proportions of these quantifying LDs in relation to all quantifying expressions (cf. '% per Q') and their proportion out of all LD tokens (cf. '% per LD').

Table 4.4: Absolute frequencies and frequencies per 100,000 words ('norm.') of the quantifying expressions *all*, *a lot*, *many*, *most* and *some*; LD tokens with these quantifiers in the initial element (absolute frequencies, proportions per total number of quantifier ('% per Q') and proportions per LD tokens ('% per LD)).

corpus	**quantifier**		**quantifying LDs**		
	N	norm.	N	% per Q	% per LD
GB	1818	901.6	3	0.17	4.17
IRE	1562	774.4	3	0.19	2.33
NZ	1986	863.5	5	0.25	5.26
CAN	1600	759.5	5	0.31	5.15
SIN	1907	938.0	9	0.47	7.69
PHI	1329	613.5	13	0.98	7.69
JAM	2056	964.0	24	1.17	14.20
IND	2500	1157.4	51	2.04	14.33
HK	2639	1108.9	38	1.44	22.35

The percentages given in the last column of the table indicate that quantifying expressions occur most frequently in LD tokens uttered by L2 English speakers, with speakers of Hong Kong English, Indian English and Jamaican English showing the highest proportions (22.35%, 14.33% and 14.20% out of all LD tokens, respectively). The proportions are much smaller in the speech of Singapore English and Philippine English speakers (7.69% each), which are thus closer in their behaviour to the L1 English speakers.

The high proportions of quantifying LDs in Hong Kong English, Indian English and Jamaican English could also be an epiphenomenon of the generally more frequent use of quantifying expression by speakers of these three varieties. As the third column

with the normalized frequencies indicates, it is especially speakers of Hong Kong English and Indian English who use quantifying expressions more frequently than the other speakers (1108.9 and 1157.4 tokens per 100,000 words, while the other frequencies per 100,000 are all below 1000 tokens).[14] Yet, if we calculate the proportions of quantifying LDs in relation to the overall number of quantifying expressions, we see that Hong Kong English, Indian English and Jamaican English still show the highest percentages (cf. fifth column). The percentages are very small indeed, but still I would argue that these findings show that the higher numbers of quantifying LDs in Indian English, Hong Kong English and Jamaican English are not an artefact of the higher frequency of use of quantifying expressions generally. Rather, these speakers really seem to include a quantifier in the initial element more frequently than the speakers of the other varieties.

As for the motivation for using such LD tokens, I can only speculate. In the Indian English data, the expression *all* accounts for a high proportion of quantifying LDs (21 items out of 51), which may speak for an additional discourse function of the LD construction in this variety of English, namely a summary function. To illustrate this function more clearly a number of further examples are given in (4.10) to (4.12). Note that this summary function has also been identified for fronting constructions and non-initial *there*-constructions in Indian English (Lange 2012: 137).

(4.10) So <,> three miles <,> that is about five kilometres everyday we used to walk in the morning and return back in the evening like that <,>
And sometimes <,> so we were nearly twenty <,> persons uh uh <,> fifth standard sixth standard seventh standard eighth standard like that <,>
So **all of us together** <,> we we were going to <,> take a room <,> uhm <,> and uh on rental basis <,> stay there
(ICE-IND:S1A-076)

(4.11) I'm </w> a <,> pure vegetarian <,>
Sometimes we take eggs <,,>
Like my family **all the people** they are taking non-veg <,,> but only me I'm the pure vegetarian one <,>
(ICE-IND:S1A-007)

(4.12) And that man he was the only son of that man <,,>
And uh he had some three or four sisters <,> **all of them** they were unmarried
(ICE-IND:S1A-069)

These examples show that LD tokens involving the quantifier *all* may be used to summarize a previously mentioned group of entities, as in (4.10), where *all* comprises the nearly twenty persons who walked to school together mentioned previously. In these sentences, the expression *all* seems to be used to emphasize that the proposition applies to really all of the people mentioned before. It seems that the expression *all* has

14 The frequency of quantifying expressions in Indian English is significantly different from all other varieties ($p < 0.001$) but Hong Kong English. The difference in frequency between Hong Kong English and the other varieties is also statistically significant (at $p < 0.01$ for Jamaican English and at $p < 0.001$ for all other varieties).

some additional meanings and functions in Indian English, which might be worth looking into in a more comprehensive and thorough way.

In the Hong Kong English data, the most common quantifiers found in LD constructions are *some* (14 tokens or 36.8% out of all quantifying LDs), *most* and *many* (10 tokens or 26.3% each). They often occur in combination with the noun *people*, that is, *some people they* or *most people they*. Possibly, Hong Kong English speakers are influenced by the classifier system of Cantonese. In Cantonese, or in Chinese dialects generally, each noun is assigned a classifier, comparable with the assignment of gender in many European languages. Two different types of classifier are usually distinguished, measure classifiers (sometimes also called measure words) and type classifiers (or simply classifiers).[15] While the former "denote quantities of an item", the latter "belong with the verb and classify it in terms of some intrinsic feature" (Matthews/Yip 1994: 109). These intrinsic features may relate to the entity's shape, natural kind or function. The following examples nicely illustrate the functioning of the classifiers (note that these examples are from Mandarin).

(4.13) *san* *zhang* *baozhi*
three CL newspaper
'three pages of newspaper'

(4.14) *san* *fen* *baozhi*
three CL newspaper
'three subscriptions of newspaper'

(4.15) *san* *jia* *baozhi*
three CL newspaper
'three newspaper agencies'

(Del Gobbo 2014: 41-42)

In all three clauses we have the items *san* 'three', followed by a classifier, which in turn is followed by *baozhi* 'newspaper'. Yet, depending on the type of classifier, we get totally different meanings: pages of the newspaper, subscriptions or agencies.

For the present study, the quantity classifier or measure word *dī* is particularly interesting. Like English *some*, it is used to denote a quantity of either countable things or uncountable substances (Matthews/Yip 1994: 115). Consider the following examples for illustration.

(4.16) dī saimānjái
'the/some children'
dī séui
'the/some water'

(Matthews/Yip 1994: 115)

15 The distinction between measure words and classifiers is not a straightforward one and the issue has been hotly debated in the literature. See, for example, Her and Hsieh (2010), Her (2012) or Del Gobbo (2014).

As indicated by the English glosses, the classifier *dī* can either mean 'the' or 'some', with consequences for the meaning of the whole clause, of course. Another interesting feature of the Cantonese classifier system is the 'bare classifier' construction which is not a general Sinitic feature but is characteristic of Yue dialects (which Cantonese belongs to) and other dialects that have been in contact with them (Bauer/ Matthews 2003: 153). In this construction the classifier serves like a determiner to specify that the noun has definite reference, as illustrated in the following example.

(4.17) *kan ok how taj*
CL house very big
'The house is very big.'

(Bauer/Matthews 2003: 153)

Given the classifier system in Chinese, then, it might be possible that Hong Kong English speakers transfer features of the Chinese noun phrase syntax onto English noun phrases. That is, they use an English expression which they think corresponds to the Chinese classifier in the given context and we thus find more quantifying expressions. Another plausible scenario, suggested by the example sentences given above, is that for Hong Kong English speakers the distinction between determiners and quantifiers is somehow blurred because of the classifier system in Cantonese.

Note, however, that these are only tentative suggestions which call for further research. It might be worth analyzing quantifying expressions in Hong Kong English in general. They might have assumed new meanings and functions in the same way as seems to be the case with the quantifier *all* in Indian English. An interesting piece of research indeed yet beyond the scope of the present study.

In the Jamaican English data, the most common quantifiers occurring in LD constructions are *some* and *a lot* (7 tokens or 29.2% each out of all quantifying LDs). Interestingly, the latter type occurs much less frequently in all the other varieties analyzed, both in LD tokens and in general. Yet, for the moment I cannot think of any reason why Jamaican English speakers seem to prefer *a lot* over *many* and *much*. The former is the most informal variant among these three quantifiers. Since the present study draws its data from informal conversations, it is not particularly surprising to find a large number of the quantifier *a lot*, but this does still not explain why it is more frequent in the Jamaican English data than in the data of the other varieties.

The final two sections discuss two specific types of LD construction. The first involves the particle *for* in the preclausal element and the second has a left-dislocated pronoun rather than a full noun.

4.1.5 *For*-LDs

This section examines a specific type of LD, which I labelled *for*-LD. This is because the preclausal element is introduced by *for*, making the LD token similar to the *as for*-construction, which is commonly used in written and spoken English to establish a topic. In what follows a number of examples are given for illustration. The extract

in (4.18) is taken from a conversation between two students who talk about student life in Japan and Hong Kong. Speaker A, probably from Japan, asks speaker B if students in Hong Kong do also have part-time jobs. Note the occurrence of three LD constructions in B's turn, two *for*-LDs and one 'normal' LD structure. The relevant sentence parts are marked in bold print.

(4.18) A: Yes do the part-time job and they they do hold parties so many times and so how about you the students in Hong Kong

B: Uhm Hong Kong students for me and also for my friends also but have also part-time jobs such as uhm uh private teacher to teacher to teach those secondary students or the primary students English or Mathematics

And apart from having the part-time job usually we'll tend to play some kind of sports

Uh **for me** I like badminton and squash and **some of them** they would like to play uhm other kinds such tennis and but usually **for girls** they like shopping

(ICE-HK:S1A-045)

The co-occurrence of 'normal' and *for*-LDs in one sentence suggests that both types of construction are well entrenched in the speaker's grammar and that they can possibly even be used interchangeably. Typically, it is subjects that are fronted in this way, yet there are also some rare cases of fronted objects, as in (4.19). In this extract, a teacher talks about a task she asked her students to do, namely writing a radio programme.

(4.19) A: I ask my students to produce the radio programs uhm which lasts about uh last for about twenty minutes [...] But I mean if you know for the radio program they include

B: Songs.

A: That's it and also I mean **for my class** I ask them to do four at least four different things.

(ICE-HK:S1A-016)

For-LD constructions usually occur in declarative sentences, as illustrated by the examples above. Yet, we also find them in questions, which further underlines the impression that they are well established and flexibly used by at least some speakers of English. The sentence in (4.20) is an example of a *for*-LD in the form of a question. The speakers talk about A's mother, who has recently retired. In addition to the *for*-LD in speaker B's question, also note the LD proper in A's response.

(4.20) B: How old is she

How old do you I will you be when you retire

I mean what's the retiring age

A: I guess I will be sixty

B: **For your mom** is is she fifty yet

A: My mom ya she's fifty-eight

(ICE-HK:S1A-059)

The most common type of initial element in *for*-LDs found in the ICE corpora involves the sequence *for me I*, as in the first example in (4.18). Additionally, we also find sequences with other pronouns, such as *for us we*, *for her she* or *for him he*, or noun phrases, such as *for girls they* in the second *for*-LD construction in example (4.18).

What all the above examples show is that structurally *for*-LDs are basically 'normal' LDs, simply with the addition of the initial *for*. They also serve the function of breaking down the utterance into smaller units, introducing the topic first, followed by some new information about the topic. Speakers often use *for*-LDs to express their opinion, frequently in contrast to some other person's opinion. The types of verb that occur most frequently in the core of the clause underline this impression. We often find verbs of emotion, such as *like*, *enjoy* and *hate*, or verbs expressing one's opinion, such as *think*, *believe* and *would rather*. In order to further underscore their point of view some speakers add the adverb *personally* to their utterance or introduce it with the conjunction *but*. Consider the following examples for illustration. The extract in (4.21) is from a conversation about job opportunities. Example (4.22) is taken from a British discussion forum for people interested in archery (Archery Interchange UK Forums), included in the Corpus of Global Web-based English (GloWbE).

(4.21) But uhm <,> as uh for me <,> uh <,> uh for <,> as an Economic student <,>
And somebody <?> ask </?> somebody said that will will uh we have uh an advantage for finding jobs such as management trainee
But in fact for me I find that <,> uh the business administration stream student is much better than us <,>
And maybe we just have some advantage in the banking field

(ICE-HK:S1A-012)

(4.22) Tabs are a very personal thing. # **Personally for me I** love the Black Widow Tab, I have tried a number but i always go back to the Black Widow

(GloWbE, GB G, Struggling to find the right tab!, accessed 24/07/2014)

The *for*-LD construction is also mentioned in the Oxford English Dictionary (OED), which gives an example of the construction dating before 1657: *For old Marinus, I know not how to excuse him.* Referring to the type *for me*, the OED gives the French *pour moi* as a possible model on which the English construction has been formed and adds that the construction is now obsolete: "The parenthetic use, as in ***for me*** = as for me, for my part (= French *pour moi*), is now obsolete".[16] This is interesting in so far as the present data prove otherwise. The structure does not seem be obsolete at all. Admittedly, it is rather rare in the L1 English varieties analyzed, but quite a number of items could be found in some of the L2 English data, as Table 4.5 shows.

What is interesting to note is that *for*-LDs seem to occur in particular in the speech of those L2 English speakers who have a topic-prominent background language. As

16 "for, prep. and conj." OED Online. Oxford University Press, September 2014; accessed 15/11/2014

can be seen, *for*-LDs are by far most frequent in Hong Kong English, followed by Singapore English and Philippine English. The high frequency of *for*-LDs in exactly these varieties is probably not owed to chance. The question is what makes these speakers use *for*-LDs more frequently than speakers of other English varieties. I would suggest that we are here dealing with a case of grammaticalization that is triggered or accelerated by language contact, that is, the development of the *for*-LD construction is an instance of contact-induced grammaticalization (Heine/Kuteva 2003, 2005, 2010). More on this below. Let us first consider the distribution and nature of the *for*-LD construction in some more detail.

Table 4.5: *For*-LDs in nine ICE corpora (absolute token frequencies).

GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
2	-	-	3	17	12	3	8	43

Hong Kong English does not only show the highest frequency of *for*-LDs but also the greatest variation with respect to the realization of the construction. While *for*-LDs typically involve a pronoun in most other varieties, Hong Kong English speakers create much more complex structures, often of the form *for* NP, as the examples in (4.19) and (4.20) illustrate; 30 items out of the 43 *for*-LD tokens are of this type. This observation indicates that the construction is well-established in Hong Kong English. This seems also to be the case for Singapore English and Philippine English because the construction is used by many different speakers in various texts. That is, rather than being an idiosyncratic feature of only a few speakers the construction seems to be widely (although rarely) used in the speech communities.[17]

In the literature, no evidence of the *for*-LD construction has been found and since the 'direct conversation' files of the ICE corpora constitute a rather small database, it was decided to search for the construction in GloWbE, the Corpus of Global Web-based English. This was also done to test the hypothesis that the construction is mainly used by L2 English speakers with a topic-prominent L1.

The investigation of GloWbE turned out to be more difficult than expected. The more complex sequences with *for* NP, which in ICE often contain some premodifying adjectives or some postmodifying material, could not comprehensively searched for. The investigation of constructions with *for me* and *for us* worked out well, but these pronominal items represent only part of the picture and do not satisfyingly reproduce the results of the ICE corpora (recall that most items in ICE-Hong Kong do not contain a pronoun). A feasible approximation to the *for* NP-construction could be achieved with the strings *for the _ they*, *for some _ they* and *for many _ they.* [18]

17 In ICE-Hong Kong, the 43 sentences are uttered by 22 speakers in 20 texts; in ICE-Singapore, the 17 *for*-LDs occur in the speech of 17 speakers in 16 texts; and in ICE-Philippines, the 12 constructions have been found in 12 different texts.

18 The interface in GloWbE allows for collocate searches. The sequences *for some _ they* and *for many _ they* were searched for with *they* as the collocate and up to five words in between *for some* or *for many* and *they*. With the sequence *for the _ they* a similar collocate search did not work out because too many items would fit that pattern. Consequently, the search was performed with the help of the

Figure 4.6 represents the findings in GloWbE and ICE differentiated according to the form of the preclausal element.[19] Making a distinction between preclausal pronouns and nouns makes sense because in the ICE data Hong Kong English sticks out in particular in that it shows a higher proportion of nominal *for*-LDs in comparison to the other varieties, as outlined above.

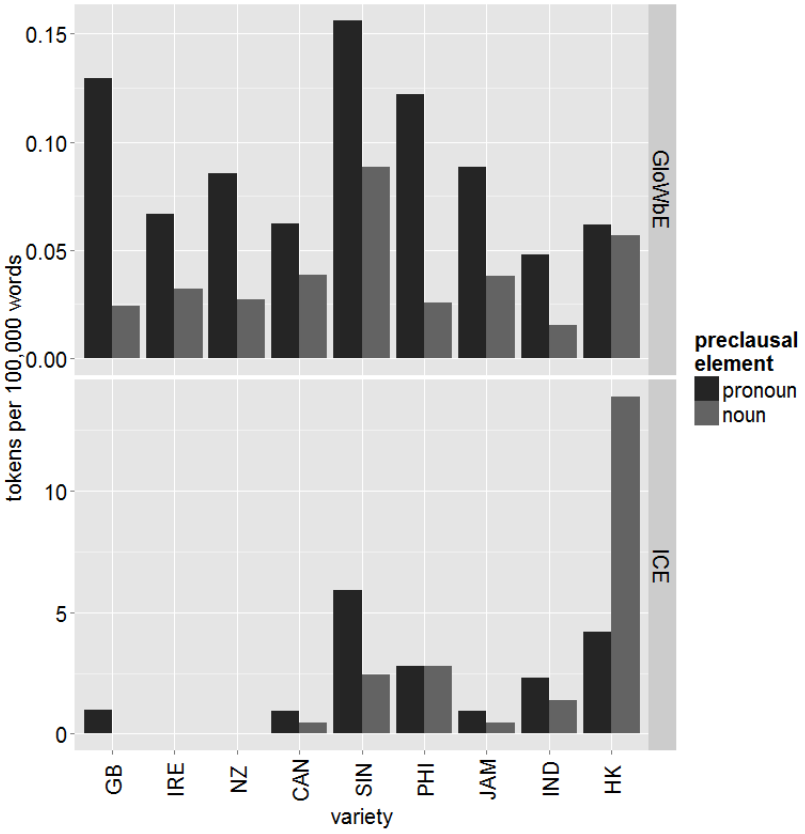

Figure 4.6: *For*-LDs with pronouns and nouns as the preclausal element in ICE and GloWbE (frequencies per 100,000 words).

As can be seen in Figure 4.6, the construction is more evenly distributed in GloWbE than in ICE. In the latter group of corpora we find a great discrepancy between the frequencies in the various data: while there are zero occurrences in ICE-Ireland and ICE-New Zealand, we find much higher numbers in ICE-Hong Kong. The findings furthermore suggest that *for*-LDs are a feature of spontaneous spoken language rather than of the language of blogs and discussion forums (as represented by GloWbE) since we find more than 18 items per 100,000 words in the ICE data (Hong Kong) but only roughly 0.25 items in GloWbE (Singapore). The frequencies in GloWbE are extremely

wildcard *, which stands for any one word. I was interested in items with two or three words in between *for the* and *they*, so the precise search syntax was *for the * * they* and *for the * * * they*. If a search returned more than 100 hits, I worked with 100-word samples.

19 The frequencies per 100,000 words in the various ICE samples and GloWbE can be found in Appendix 6.6.5.

low and it is therefore difficult to draw reliable conclusions. Yet, some interesting trends can still be observed, which are definitely worth reporting.

Turning to the distinction between the pronominal and the nominal version of the *for*-LD construction, it is interesting to note that in Hong Kong English the nominal version is the more common type while we find more items of the pronominal type in all the other English varieties (in both GloWbE and ICE). Furthermore, among the data obtained from GloWbE the *for* NP-construction is most frequent in Singapore English and Hong Kong English, exactly the two English varieties which are in contact with Chinese dialects.

Summing up the results, it can be said that the pronominal *for*-LD construction is used by speakers of L1 and L2 English varieties alike to express their opinion and feelings, often in contrast to some other person's opinion. This type of *for*-LD has to be distinguished from the more complex NP version, which figures prominently in the two English varieties with a Chinese background language. It seems that speakers of these two varieties of English have widened the scope of the *for*-LD construction extending it to more complex subjects (and objects, occasionally). This suggests that the preposition *for* has grammaticalized into a flexible and well-entrenched topic marker in Singapore English and Hong Kong English. Grammaticalization is here understood as the process "whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions" (Hopper/Traugott 2003: 18; on the universal principles of grammaticalization also see Heine et al. 1991; Bybee et al. 1994). In the present example, the preposition *for* develops the new grammatical function of a topic marker, triggered by the attempt of the English speakers to structure their sentences according to the principles of their topic-prominent L1s. Since the process of grammaticalization is here triggered or accelerated through language contact, we may also speak of contact-induced grammaticalization (Heine/Kuteva 2003, 2005, 2010). What is interesting in the present case is that the L2 English speakers make use of their full repertoire of linguistic features to create the *for*-LD construction: they aim at structuring their utterances in a topic-prominent way, which they are familiar with from their L1s, but they choose English constructions as the models for the new structure. The L2 English speakers recognize that in English - a subject-prominent language - a sentence can be structured around a topic by means of the *as for*-construction (e.g. *As for books, I like Edgar Allan Poe*). The functional similarity between LD and the *as for*-construction is also noted in the literature, with Lambrecht (1994), for example, writing that "the *as-for* construction is […] a subtype of the detachment or dislocation construction" (1994: 152; also cf. Gundel 1988). Leaving out the particle *as*, the *for*-LD construction is very likely based on the model of the *as for*-construction taking over the function of topic establishment.

Another construction which I would suggest impacts on the development of the *for*-LD is the *for NP*-structure, which commonly occurs in clause-initial position and serves to express an opinion or to establish a topic, as the sentences in (4.23) illustrate. Recall that *for*-LD constructions typically serve precisely these functions.

(4.23) a. I told you I told you I told you **for me** it's the preschoolers **for you** it's the grade six (ICE-PHI:S1A-082)

b. So he just says it casually but <,> **for me** it was terrifying that he would <,> go to jail and never come back (ICE-CAN:S1A-043)

c. **For them** it is very convenient if it is in Delhi <,> (ICE-IND:S1A-070)

Structurally, these sentences are very similar to the *for*-LD construction. The only difference is that the subject of the main clause is co-referential with the element following *for* in *for*-LDs but not in *for NP*-structures. In (a), for example, we have *me* and *you* after *for*, but the subject is *it*. The structures also differ semantically, as the German translations clearly indicate. In the sentences in (4.23), *for* is translated as 'für', for example, 'für mich war es Angst erregend, dass er ins Gefängnis gehen würde' as a possible translation of the *for*-structure in (b). In *for*-LDs, on the other hand, *for* has lost this meaning. For example, the sentence *for girls they like shopping* in (4.18) would not be translated into 'für Mädchen sie lieben Shopping', but you would rather say 'Mädchen, die lieben Shopping'.

The process of contact-induced grammaticalization, then, appears to be a plausible pattern for the development of the *for*-LD construction, as it shows many of the main characteristics of the process of grammaticalization: it is triggered (or accelerated) through contact with other languages; *for* has lost its original meaning and has acquired a more abstract meaning (desemanticization); it has assumed the new function of topic marking (extension); the original form of *for* and that used as a topic marker coexist; the meanings or functions of the *as for*-construction and the *for NP*-construction constrain the functions of the *for*-LD, that is, it is mainly used to establish a topic and voice an opinion (cf. Hopper/Traugott 2003: 2; Heine/Kuteva 2005: 15, 80).[20] In Hong Kong English, evidence in support of extension also comes from the fact that the *for*-LD construction allows for elements to occur after *for* other than pronouns, while the pronominal variant seems to be the more common type in the *for NP*-structure and the speech of most other speakers.

What the discussion of the development of the *for*-LD construction above has shown is that we are here dealing with a rather complex process of grammaticalization. There is not one single language that is the source of the new structure, but rather we have several languages interacting to create the *for*-LD structure. Such linguistic creativity, however, can be expected in the case of bi/multilingual speakers, who, in the attempt to adhere to the principles and expectations of the language-particular setting, make use of the full repertoire of linguistic features at their disposal (Matras/Sakel 2007: 852). Furthermore, the present example is interesting in that a rather abstract concept of the speakers' L1s serves as the triggering or accelerating factor, namely topic prominence. This is probably not the typical path of grammaticalization or pattern replication, but it is described as a possible scenario in the literature: "in order to replicate model language constructions, speakers may also select abstract

20 Heine and Kuteva (2010) describe grammaticalization as a subcategory of grammatical replication. A similar model is provided by Matras and Sakel (2007), who see grammaticalization as a subtype of what they call 'pattern replication'. Since the focus of the present study lies on information-packaging constructions, the process of grammaticalization as understood in these different models will not be discussed in detail here. The interested reader is referred to these works for more information on the theoretical assumptions of the different models and the similarities and differences between them.

morphosyntactic operations as pivots" (Matras/Sakel 2007).[21]

A difficult question, which is raised by the present analysis but which it is unable to answer conclusively, concerns the precise role of language contact in the process of grammaticalization. That is, it is difficult to say whether language contact has actually triggered the grammaticalization process or whether it is 'only' speeding up a change that would have happened anyway. Recall that the *for*-LD construction had already existed in Early Modern English, but was then abandoned, according to the OED. Hence, the construction is probably not unfamiliar to speakers of English. It might also be the case that we are here dealing with a general development spearheaded by certain varieties of English. That is, while *for*-LDs are more prevalent in the speech of English speakers with topic-prominent L1s in the present data, it might well spread across varieties of English around the world in the future. Support for this assumption comes from a similar development that has been described for another topic marking structure in English, namely the *as far as*-construction, where the verbal coda *goes* or *is concerned* is increasingly more often deleted (Rickford et al. 1995; Britain 2000). Consider the following example, where we would expect that the final sequence of words be *as far as bills and savings* ***are concerned***.

(4.24) I've decided that I'm just going to work with what I earn from teaching and just work with that as far as um bills and savings

(ICE-NZ:S1A-085)

What the *as far as*-construction without verbal coda and the *for*-LD have in common is that they are both used to mark a topic and both have developed out of more complex constructions by leaving out certain elements (*goes/is concerned* and *as*, respectively). The constructions differ, however, in that the *for*-LD seems to be a phenomenon largely restricted to the speech of L2 English speakers with a topic-prominent background language while the deletion of the verbal coda in *as far as*-constructions has so far been attested for American English and New Zealand English.

It is clearly necessary to investigate larger sets of spoken data or more varieties of English (as provided by GloWbE) to test whether the grammaticalization of *for* into a topic marker is indeed largely restricted to L2 English varieties with topic-prominent background languages or whether these varieties are in the lead of a general development. Additionally, it might be interesting to compare the results with constructions introduced by *with me*, *to me* and *about me*, which seem to function in a similar way, as the sentences in (4.25) illustrate.

(4.25) a. **With me** I lost my mother when I was only seven

(GloWbE, NG B, Interview with Asa Asika… 9ja Breed)

b. **To me** I'm not a punctual person (ICE-IRE:S1A-061)

21 Matras and Sakel (2007) describe the process that is responsible for pattern replication as 'pivot-matching', which "involves identifying a structure that plays a pivotal role in the model construction, and matching it with a structure in the replica language, to which a similar, pivotal role is assigned in a new, replica construction" (2007: 830). In the present case, the pivot in the model language is the topic-prominent sentences structure, the matching structures in English, the replica language, are the *as for*-construction and the *for NP*-structure.

c. **About me** I am a BS Mathematics student at UPLB
(GloWbE, PH G, The Mania That Was - Sporty Guy)

The initial expressions *with me*, *to me* and *about me* seem to be used when the speakers want to say something about themselves, possibly in comparison to some other person. Research in this direction will certainly be worth the effort.

4.1.6 Pronominal LDs

Another specific type of LD construction that shows some variation across the varieties analyzed involves a pronoun as the preclausal element (without *for*). Examples are given in (4.26) and (4.27).

(4.26) **Me I** am a Development Policy student (ICE-PHI:S1A-072)

(4.27) **Us we** make good music (ICE-PHI:S1A-063)

Lambrecht (1994: 183) describes this type of construction as frequent and he argues that this is because "the left detachment construction is often used to mark a shift in attention from one to another of the two or more already active topic referents". Interestingly, in the ICE data analyzed, the pronominal LD construction has been found in a substantial number only in ICE-Philippines, where it makes up more than 12% of all LD tokens (21 instances out of 169 LD tokens, dispersed over 14 texts and spoken by 16 different speakers).[22] In the other corpora the construction is very rare, with one single example each occurring in ICE-Jamaica, ICE-Canada and ICE-Great Britain.

Most of the tokens attested for ICE-Philippines are of the form *me I*; two examples are of the form *us we*. This suggests that the pronominal LD construction is mainly used when the speakers want to say something about themselves. They introduce themselves as the topic of the following message expressing their opinion, often in contrast to some other person's opinion. This observation agrees with Lambrecht's (1994) claim that pronominal LDs "often have a 'contrastive' function" (1994: 183).

Mesthrie (1992: 116) also finds this subtype of the LD construction in his South African Indian English data and describes it as a "minor variant". Huber and Dako (2004) and Huber (2014; presentation at ICAME35) report that pronominal LDs are more frequent in Ghana English than in British English. Given these findings and comments by different authors, it seems that the pronominal LD construction is not unique to Philippine English, as the results of the present study suggest.

Especially intriguing is the discrepancy between the findings of the present study with respect to British English (one example) and Lambrecht's (1994: 183) claim that

22 Recall that in the analysis of the information status and topic persistence of the initial elements in LD tokens in section 4.1.2 it was noted that ICE-Philippines stands out among the other varieties analyzed in that it shows a high number of known initial elements that are pronominally referred to in the following five sentences. To a great extent, this is due to the high frequency of pronominal LDs in this variety of English.

pronominal LD tokens are frequent. This may be due to the make-up of the ICE component, that is, the somewhat higher level of formality of the 'direct conversation' files included in ICE-Great Britain. To test this hypothesis the distribution of LD constructions with the initial sequence *me I* was analyzed in GloWbE.[23] It turned out that pronominal LDs of this type are very rare in GloWbE. Among the nine varieties of English which are subject to the present study, British English shows the highest frequency with less than 0.07 items per 100,000 words, a very small frequency indeed.[24] Out of interest and because of Huber's (2014) and Huber and Dako's (2004) findings in Ghanaian English the GloWbE components representing Ghanaian English and Nigerian English have also been investigated in the present analysis. Interestingly, both African varieties of English show higher frequencies than British English, but the frequencies are also still very small (0.136 items per 100,000 words in Ghanaian English and 0.350 items in Nigerian English).

In sum, then, the frequencies of *me I*-LDs are so small in all varieties of English analyzed that it is difficult to draw reliable conclusions from these results. Although the British English data in GloWbE show the third highest frequency after the two African varieties of English, I am reluctant to say that this confirms Lambrecht's claim because the frequency is so very low. What the analyses of *for*-LD constructions and pronominal LD tokens with initial *me I* show, however, is that the comparison of ICE and GloWbE is a valuable research set-up, but it also has its limitations. The phenomena under consideration are very infrequent in GloWbE - despite the huge size of the corpora - which suggests that they are clearly features of informal conversations and that the language of blogs and discussion forums is still quite dissimilar from spontaneous spoken discourse.

4.1.7 Summary

One major issue that has been discussed in the present section involves the high frequency of LD constructions in the speech of Indian English speakers. Addressing the question of what motivates Indian English speakers to use LD tokens so much more frequently than the speakers of the other varieties, a number of pragmatic and syntactic features has been examined. The creation of topic continuity has been reported in the literature as a major motivating factor for Indian English speakers to use the structure (Lange 2012). However, this claim cannot be confirmed in the present study even with the interpretation of the notion 'topic continuity' being taken somewhat

23 Since the search for the sequence *me I* in GloWbE returned many items that were not instances of LD tokens, I restricted my analysis to 100-word samples to extract all pronominal LDs. This was a feasible amount of data to go through and an efficient way to get an impression of the distribution of LDs with *me I* across varieties of English in GloWbE. The raw frequencies were then normalized to frequencies per 100,000 words. This was necessary, first, to get results that are comparable with those obtained from the ICE samples and, second, to make the results within GloWbE comparable because the subcorpora included in GloWbE consist of very different amounts of texts and words (see the section on data and methodology and Appendix 6.6.5 for more information on the make-up of GloWbE).

24 The frequencies of LDs with initial *me I* in GloWbE are given in Appendix 6.6.5.

wider. It seems that speakers of L2 English varieties generally use LD constructions whose preclausal elements refer to old information more often than the speakers of the L1 varieties analyzed, thus creating topic continuity. For the L1 English speakers, on the other hand, the simplifying function seems to play a more important role than for the L2 English speakers.

Additionally, two features of the preclausal element in LD constructions have been analyzed. First, its syntactic function, or rather that of the co-referential pronoun, has been examined hypothesizing that the high frequency of LD tokens in Indian English might be due to a more flexible usage of the construction in this respect. Yet, it turns out that Indian English speakers are least flexible among the varieties analyzed, showing the highest proportion of the 'default' subject LD. Irish English, on the other hand, shows some more variation, with the somewhat higher incidence of possessive LDs possibly being due to influence from the Irish background language.

The second feature examined with respect to the preclausal element is concerned with the presence of a quantifying expression. Speakers of Hong Kong English, Indian English and Jamaican English turn out to use 'quantifying LDs' more frequently than the speakers of the other varieties. In Indian English, the quantifier *all* is particularly common. It seems to have acquired new meanings and functions in this variety of English. Hong Kong English speakers might be influenced by the classifier system of their Chinese background language.

Finally, two specific types of LD token have been discussed, *for*-LDs and pronominal LDs. In *for*-LDs, the preposition *for* has grammaticalized into a topic marker, influenced by the topic-prominent background languages of some speakers and the English *as for-* and *for NP*-constructions. English speakers with topic-prominent background languages seem to use the construction more frequently than speakers of the other varieties because they are more sensitive to the notion of topic and therefore tend to structure their sentences according to the principles they are familiar with from their L1. Because of the triggering or accelerating function of the topic-prominent L1s we may speak of a case of contact-induced grammaticalization.

Pronominal LDs are almost exclusive to Philippine English in the present study, which is surprising because this type of LD has been reported in the literature to be frequent in other varieties as well.

Coming back to the question of what motivates Indian English speakers to use LD tokens so much more frequently than the speakers of the other varieties, the present analysis of pragmatic and syntactic features cannot provide any profound answers. It rather seems that the construction is simply not as marked in Indian English as it is, for example, in British English. This might be due to influence form the background languages. In Malayalam, for example, it is sufficient for a sentence constituent to be marked as the topic by placing it in sentence-initial position, but a more explicit marker of topic is also possible. In such cases, "the topic remains in first place in the sentence but is followed by a reinforcing element" which is attached to the topic (Asher/Kumari 1997: 184). Similarly, in Hindi the particle *to* may be used to mark the topic of a sentence. Possibly, Indian English speakers simply mark the subject as the topic by using the co-referential pronoun as a "reinforcing element", a strategy they are familiar with from their L1. I would thus suggest that different forces interact to unmark the LD construction and increase its usage in Indian English. For one, the

feature pool of Indian English speakers contains the possibility of marking topics explicitly by means of an ending that is attached to the topic or by means of a particle. Additionally, they are familiar with the left dislocation construction as a means of establishing a topic. Since morphological marking is not admissible in English, Indian English speakers do not directly transfer the feature from their L1 into English, but a construction that is already there assumes this function, that is, the co-referential pronoun is used as an explicit topic marker. The construction may therefore be used more frequently and it may be less marked than in, for example, British English.

4.2 Right dislocation

This section examines the nature and functions of right dislocation constructions (RDs for short). The general overview at the very beginning of this chapter has already indicated that RDs are rather rarely used by the English speakers analyzed in the present study.[25] Some of the frequencies are in fact so small that the results reported below are not always reliable and therefore must be considered with care. The structure of the section is as follows. First, the overall frequencies across the nine varieties of English analyzed will be presented. This will be followed by the analysis of a number of syntactic features. Finally, expanded forms of RD will be discussed. In contrast to the 'canonical' RD, such expanded RDs contain an operator in the dislocated element.

4.2.1 Overall distribution

As noted earlier, the group of RD tokens included in the present study comprises RD 'proper', that is, sentences with a dislocated noun phrase, as in (a), demonstrative pronoun, as in (b), or clause, as in (c), and expanded RDs, which have an operator in the detached element, as in (d) and (e).

(4.28) a. He's brilliant **your dad**. (ICE-GB:S1A-042)

b. God that would scare a kid **that** (ICE-NZ:S1A-043)

c. And that seemed to be souldestroying **being on the dole** (ICE-NZ:S1A-012)

d. He's going to Urban Dance Squad **Phil Alexander is** (ICE-GB:S1A-100)

e. He's a real pet **so he is** (ICE-IRE:S1A-049)

25 The terms 'frequent' and 'rare' are, of course, relative terms. While RD constructions are rare in comparison to most of the other information-packaging constructions analyzed in the present study, they are described as a "prominent" feature of the 5-million-word Cambridge and Nottingham Corpus of Discourse in English (CANCODE) by Carter, Hughes and McCarthy (1998). Furthermore, Cullen and Kuo (2007) point out that RDs are twice as frequent as *ought to* or the *get* passive in the spoken component of the Longman Corpus of Spoken and Written English.

Figure 4.7 presents the frequencies per 100,000 words of RD tokens in the nine varieties of English analyzed. Interestingly, while it is the L2 English varieties that show the highest frequencies of LD tokens it is now the L1 English varieties for which higher frequencies can be attested. An exception in this respect is Canadian English, with the frequency of RD tokens being smaller than those of Indian English and Singapore English (16.6, 26.9 and 23.6 tokens per 100,000 words, respectively).

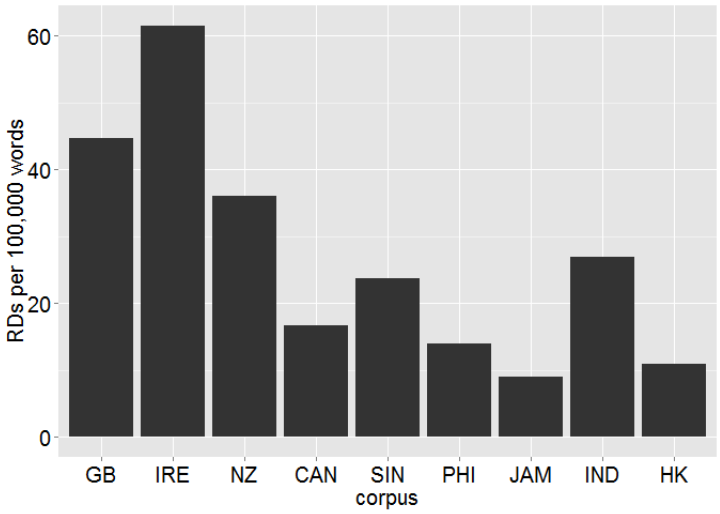

Figure 4.7: Right dislocation in nine ICE corpora (frequencies per 100,000 words).

Irish English shows the by far highest frequency (61.5 tokens per 100,000 words). This is mainly due to the large number of expanded RDs that can be attested for this variety of English. More on this expanded form below. The differences in frequency between Irish English and the L2 English varieties are statistically highly significant ($p < 0.001$); the same holds for Canadian English. For New Zealand English it is significant at $p < 0.05$ and for British English the difference is not significant ($p > 0.05$).

Among the L2 varieties, Indian English and Singapore English show the highest frequencies of RD tokens.[26] The somewhat higher frequency in the former variety is possibly due to an overuse of RD tokens with right-dislocated demonstrative pronouns in comparison to the other varieties. Singapore English shows a substantial number of tokens in which the co-referential pronoun is omitted. These items will be discussed in the following subsection, which deals with one of the major functions of RD tokens, namely the emotive or evaluating function.

26 The difference between these two varieties is not statistically significant but that between Indian English and Philippine English ($p < 0.05$), Jamaican English and Hong Kong English (both at $p < 0.01$).

4.2.2 Emotive or evaluating function

As noted earlier (cf. section 3.2.2), RD constructions often have an emotive dimension, that is, speakers often use them to express their feelings, opinions and attitudes. Such tokens usually serve the function of bonding with the interlocutor. In most of the RD tokens that are associated with some form of evaluation we find an evaluative adjective in the main clause, for example, *brilliant*, *good*, *great*, *funny*, *silly*, *difficult* etc.; or there is a noun phrase that is clearly evaluative, such as *bastard*, *swank* or *rudey*, as in the following sentences.

(4.29) a. Yeah he's a <,> bastard so he is (ICE-IRE:S1A-014)

b. He was a great swank yesterday at Mass so he was (ICE-IRE:S1A-067)

c. She's a rudey that girl (ICE-NZ:S1A-036)

In addition to evaluative adjectives and nouns, we also find some rare instances of RD tokens in which the verb conveys some form of evaluation, as in *God that would scare a kid that* (ICE-NZ:S1A-043). Looking at all the RD tokens that are associated with the expression of feelings and attitudes, it turns out that Canadian English differs from the other L1 English varieties in this respect, as can be seen in Table 4.6.

Table 4.6: RD tokens that serve an emotive or evaluating function (absolute frequencies of 'emotive RDs' and percentages out of all RD tokens).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
RD tokens	90	124	83	35	48	30	19	58	26
emotive RDs	48	62	38	12	36	12	8	21	11
emotive RDs %	53.3	50.0	45.8	34.3	75.0	46.7	42.1	36.2	42.3

While we find high proportions of such RD tokens in British English (53% of all RD tokens), Irish English (50%) and New Zealand English (46%), in Canadian English the proportion is much lower, accounting for only 34% of all RD tokens. The proportion is in fact the smallest among all varieties analyzed. From this we may conclude that Canadian English speakers stick out in two respects: first, among the L1 English speakers they show the by far smallest frequency of RD tokens and, second, if they use the construction, it less often serves an emotive or evaluating function.

As can be seen in Table 4.6, the highest proportion of RD tokens serving an emotive or evaluating function can be attested for Singapore English (75%). A substantial proportion of these emotive RDs in turn is made up by a specific type of RD that has a very succinct form, as illustrated in the examples in (4.30) to (4.32).

(4.30) B: Ya and that ah you see in the book does it talk about the upper class and the lower class of society hor

A: Oh yes

Oh yes

B: **So pretentious ah the higher class**

(ICE-SIN:S1A-090)

(4.31) A: Ya I mean it's so huge you know the print
Look at this
C: So they're pasted on uh?
A: **Not so huge uh this one**
B: Because this is from imported from Indonesia

(ICE-SIN:S1A-008)

(4.32) A: Cannot when I when I wash my hair
Aiyah I don't want to go out **dirty my hair**

(ICE-SIN:S1A-007)

Note that in all of these instances the co-referential pronoun and the copular verb *be* are missing. The omission of *be* is a common feature of many L2 and vernacular varieties of English, including Singapore English (e.g. Kandiah 1998). A closer inspection of the items suggests that they can equally well be understood as fronting constructions and that their classification as RD tokens is debatable. Compare the following 'translations' into 'full' RD tokens and fronting constructions: *It is so pretentious, the higher class* versus *So pretentious the higher class is*; *It is not so huge, this one* versus *Not so huge this one is*; and *It is dirty, my hair* versus *Dirty my hair is.* These unclear cases are difficult to disambiguate because either way constitutes a plausible interpretation. Yet, the context can help to identify the function of the constructions and thus disambiguate the sentences. As noted earlier, RD tokens mainly serve an emotive or clarifying function. Fronting constructions, on the other hand, are often used to put emphasis on the fronted element or to draw a link to the preceding discourse. Furthermore, the right-dislocated element in RD tokens always contains discourse-old information, which is not necessarily the case for fronting constructions.[27]

Looking at the context, then, examples (4.30) and (4.31) clearly turn out to be RDs rather than fronting constructions. The example in (4.32), on the other hand, is more difficult to disambiguate. The conversation in (4.30) is about the upper and the lower classes of the society. After having said *so pretentious*, speaker B realizes that it is not quite clear who she is referring to and she therefore adds, after a filled pause (cf. *ah*), that she means the upper class. Admittedly, the adjective *pretentious* could also occur in initial position to be emphasized, which would speak in favour of interpreting it as a fronting construction. Yet, since the emotive function is so common with RD tokens and since fronted adjectives occur rarely across varieties of English (cf. the analysis of the preposed elements in fronting constructions in section 4.3) the interpretation as a RD construction seems more plausible. Listening to the tape recordings would further help to disambiguate the sentence because if we were dealing with a fronting construction there would be more stress on *so pretentious* than if they were RD constructions. With RDs we would find some more stress on *the upper class* than would

27 More information on fronting constructions are given in sections 3.2.3 and 4.3.

be the case with fronting constructions.

The example in (4.31) is also quite likely an instance of a RD token since speaker A uses a very similar construction, a clear example of a RD token, in her first turn (*it's so huge you know the print*). Hence, the shorter version *not so huge uh this one* is sufficient in the second turn because it is already clear what she is talking about and she is probably also pointing at the entity (cf. deictic *this*).

The example in (4.32) is more difficult to disambiguate through the context. Here, both interpretations work well in my opinion. Speaker A says that she cannot go out that night because her hair is dirty. It could well be the case that the adjective *dirty* is fronted to be emphasized. Clarification of what is dirty seems not to be so important in this case, but it is still likely.

Further evidence of treating such examples as cases of RD comes from previous research on dislocation. Lambrecht (2001a: 1057), for example, gives examples of dislocation with "null-instantiated pronominals" from several Asian and European languages. And Aijmer (1989), in her analysis of left and right dislocation in the London-Lund Corpus of Spoken English, notes that

> the speaker can utter the predication without mentioning the subject (or referring to it with a pronoun) and then add an identifying or evaluative or descriptive noun phrase in the position of the Tail (ellipsis). This is possible because the context is familiar and the hearer can be expected to know what the speaker is referring to. (Aijmer 1989: 159)

This shows that leaving out the co-referential pronoun is not uncommon. Interestingly, however, these structures seem to be particularly frequent in Singapore English (15 tokens out of 49 RD tokens). Usually, the sentence begins with an adjective phrase (cf. examples (4.30) to(4.32)), but sometimes also with a verb phrase or a noun phrase, as illustrated in (4.33).

(4.33) a. Taste like a rock you know this bread (ICE-SIN:S1A-080)

b. Big problem you know that one (ICE-SIN:S1A-035)

Such sentences with zero subjects can also be found in the other varieties analyzed, but they are less frequent: eight instances can be attested for New Zealand English, four for Irish English, two four British English and one item each for Jamaican and Hong Kong English. Additionally, it seems that in these varieties the sentences tend to be of a slightly different kind. While we usually find dislocated noun phrases in the Singapore English sentences with zero subjects, it is demonstratives that occur most frequently in the sentences of the other varieties (e.g. *Oh very famous picture that* <,>; ICE-GB:S1A-049).

The tendency to place adjective phrases in initial position and to leave out the subject pronoun may well be due to influence from Mandarin. In Mandarin, like in all Chinese dialects, pronouns that are co-referential with the (discourse) topic are often omitted in the ensuing discourse because the interlocutors already know who or what the conversation is about ("topic chaining" cf. Huang 1984; also Ritchie 1986; Matthews/Yip 1994). Ritchie (1986) argues that this feature of the Chinese language may have left traces in the speech of Singapore English speakers. Another piece of evidence comes from Lambrecht (2001a), who gives the following example from Mandarin and its English equivalent when discussing dislocation constructions with zero co-referential pronouns.

(4.34) *té nán zhǎo, wǒ zhèige*
very difficult find my this
'It is very difficult to find, this (thing) of mine.'

(Lambrecht 2001a: 1057)

This example is interesting in the current context in so far as it illustrates that it is quite common in Mandarin to put adjective phrases at the beginning of a clause. Furthermore, the Mandarin sentence is translated as a RD construction in English. This gives further support to the idea of interpreting the ambiguous examples above as instances of RD. Similar examples can be found in Cantonese, which are incidentally called right dislocation by Matthews and Yip (1994). Consider the following sentences.

(4.35) *chǐsin gàh, léih!*
crazy PRT you
'You're crazy!'

(4.36) *géi leng wo, dǐ sǎam.*
quite nice PRT CL clothes
'Not bad, those clothes.'

(Matthews/Yip 1994: 71)

The authors note that this type of word order (the adjective phrase comes first followed by a noun or pronoun functioning as the subject) is typically used in exclamations, which fits the observation that most of these items constitute emotive RDs in the present study, as mentioned above.

4.2.3 Syntactic function of the dislocated element

The right-dislocated constituents in RD tokens, or rather their co-referential pronouns, can have different syntactic functions. In the present study these include subjects, objects, complements and possessives. For illustration of these functions consider the sentences in (4.37).

(4.37) a. [They]'re nice actually **those things** (ICE-IRE:S1A-070)
b. Where did you uhm learn to do [this] then **palm-read** (ICE-GB:S1A-091)
c. I don't know what you call [it] **this side** <,> (ICE-IND:S1A-007)
d. What is [his] name **the ELT centre Professor** (ICE-IND:S1A-019)

In most cases we find co-referential subject pronouns in the core of the clause in all nine varieties of English analyzed (cf. Table 4.7), a finding which is in line with previous studies (e.g. Grosz/Ziv 1998: 298).

Interestingly, the three L1 English varieties British English, Irish English and New

Zealand English show the highest proportions of subject RDs, which make up more than 91% out of all RD tokens. And note that it is exactly these three varieties that have the highest token frequencies. Recall that with left dislocation and fronting constructions we find a similar picture (cf. sections 4.1 and 4.3): the English varieties with the highest token frequencies show the highest proportion of the default cases (subject and object, respectively). Recall also that for LD tokens a correlation between the syntactic function and the LD construction's major discourse function, namely topic establishment or the marking of a topic as salient in the discourse, can be identified. The high frequency of subject LDs can plausibly be accounted for by the fact that subjects are the prime candidates for the position of topic. With RD tokens we find a similar correlation between syntactic function and discourse function. Note, however, that the token frequencies of object RDs are fairly low (ranging between 1 and 9 tokens) and consequently the percentages are not very robust. That is, one token more or less can heavily impact on the overall proportions. Keeping this in mind, the following should merely be seen as suggestions based on the tendencies that can be gleaned from the ICE data. Further research based on larger datasets is clearly needed.

Table 4.7: The distribution of RD tokens according to the syntactic function of the dislocated element (absolute token frequencies and percentage out of all RD tokens).

corpus	**RD tokens**	**subject**		**object**		**other***	
	N	N	%	N	%	N	%
GB	90	82	91.1	7	7.8	1	1.1
IRE	124	115	92.7	8	6.5	1	0.8
NZ	83	77	92.8	6	7.2	0	0.0
CAN	35	28	80.0	6	17.1	1	2.9
SIN	48	43	89.6	5	10.4	0	0.0
PHI	30	23	76.7	7	23.3	0	0.0
JAM	19	17	89.5	1	5.3	1	5.3
IND	58	47	81.0	9	15.5	2	3.4
HK	26	20	76.9	4	15.4	2	7.7

* The category 'other' comprises complements and possessives.

In RD tokens that serve an emotive or evaluating function, speakers typically use the copular construction 'X is AP' to evaluate some entity X; and X - obviously the subject of the clause - is then further specified in the dislocated element. RDs that serve a clarifying function, on the other hand, are more flexible and work just as well with objects, complements or possessives, as the following examples illustrate. In (4.38), speaker E is asked about her favourite books. Note that speaker D's *I'm reading it* needs disambiguation because *Treasure Island* and *The Wind in the Willows* have both been mentioned right before.

(4.38) A: Can you give us an example?

E: Well yes I mean

B: Treasure Island

E: Yes well

D: The Wind in the Willows

I'm reading [it] **Treasure Island** at the moment to my son

(ICE-GB:S1A-013)

The following example illustrates a RD construction formed around an object complement. Without adding *my sister* the clause *I call her that* would remain vague. Using a RD construction in this case also allows the speaker to put more emphasis on the complement.

(4.39) A: Is he married your sister?

B: Yeah

My sister I have

No I haven't any sister

I'm the only child in family

But uhm my uncle's daughter I call her [that] **my sister**

(ICE-HK:S1A-056)

Only two possessive RD tokens can be attested for the nine ICE corpora analyzed. One has already been quoted in (4.37); it is repeated in (4.40) with some more context. In (4.40), the speakers are trying to remember the name of a professor from the ELT (English Language Teaching) centre. The person is here mentioned for the first time in the discourse, so the use of the pronoun *his* is rather surprising. It seems that speaker B is here speaking while having a picture of the professor in mind. That is why she uses the pronoun first.

(4.40) A: I was working in a ELTR na

B: Ha yes and <,> uh what is [his] name **the ELT centre Professor**?

A: Sumat Roy, Sumat Roy <,,> He was the director

B: Yeah but <,,>

(ICE-IND:S1A-019)

The other possessive RD is from ICE-Jamaica. The following sentences constitute the first few turns in the file and, consequently, it is not possible to say whether the two speakers have been talking about Amil before. It seems, however, that the person is familiar to both speakers and that A even expected B to know who he is talking about. Maybe A uses a RD construction as an answer to B's question *Whose wife?* in order to express surprise (because of B's lack of understanding) and to put emphasis on *his* and Amil.

(4.41) A: It was just from the net and they were just

His wife wrote it

B: Whose wife?
A: [His] wife **Amil** Amil <,>
B: That fellow's
Okay
A: Yeah yes

(ICE-JAM:S1A-009)

In sum, this suggests that if there is a high proportion of RD tokens that serve an emotive or evaluating function, then there is quite likely also a high proportion of subject RDs. Note that this argument is unidirectional. A small proportion of emotive RDs does not automatically entail a small proportion of subject RDs because clarifying RDs can equally well be formed around subjects. However, the PROBABILITY to have a syntactic function other than the subject seems higher with RD tokens serving a clarifying function or some other (minor and specific) function, as can be seen in examples (4.39) to (4.41).

4.2.4 Realization of the dislocated element

Across the nine varieties of English analyzed there is also some variation in terms of the realization of the right-dislocated constituent. The variants that are part of the present study include nouns, pronouns (personal and demonstrative) and clauses (finite and non-finite). Numerous examples of these types have already been given throughout the previous subsections, see, for instance, the sentences in (4.28). The distribution of RD tokens according to these realizational variants is presented in Table 4.8 (absolute token frequencies and percentages of all RD tokens).

Table 4.8: Realization of the dislocated element (absolute token frequencies and percentages of all RD tokens).

corpus	noun		pronoun		other*	
	N	%	N	%	N	%
GB	52	57.8	24	26.7	14	15.6
IRE	43	34.7	77	62.1	4	3.2
NZ	73	88.0	8	9.6	2	2.4
CAN	30	85.7	2	5.7	3	8.6
SIN	43	89.6	4	8.3	1	2.1
PHI	23	76.7	3	10.0	4	13.3
JAM	11	57.9	4	21.1	4	21.1
IND	35	60.3	18	31.0	5	8.6
HK	22	84.6	4	15.4	0	0.0

*The category 'other' contains clauses and one PP.

As can be seen, we typically find right-dislocated noun phrases that are coreferential with a pronoun in the core of the clause, which ties in with previous research: "Most commonly, a tail [i.e. right-dislocated element] consists of a full noun phrase which clarifies or repeats the referent of a pronoun in the clause that comes before it [...]" (Carter/McCarthy 2006: 194). An exception in this respect is Irish English. Here, it is dislocated pronouns that occur most frequently (62.1% of all RD tokens), which is due to the high number of expanded forms of RD because the latter almost always contain a pronoun. Among the other varieties, Singapore English shows the highest proportion of dislocated noun phrases (89.6%), while British English and Jamaican English show the smallest proportions with nearly 58% of dislocated noun phrases. Note, however, that the token frequencies in ICE-Jamaica are fairly small and therefore the percentages should only cautiously be regarded as representative (4 pronominal and 4 clausal RDs out of 19 RD tokens).

As indicated in the table, the category 'other' contains mainly clauses but also one example of a dislocated prepositional phrase. This rather atypical sentence, provided in (4.42), is from Philippine English. The conversation is about B's father, who was a busy lawyer always travelling around the country.

(4.42) B: So we never really resented the fact that he wasn't around
And when he was around it would always be fun
Uhm meal times were great you know
[...]
A: That was your special time right **during meal time**

(ICE-PHI:S1A-010)

After saying that meal times were great in her family, speaker B continues to tell A why they were so great, what they did and what they talked about (omitted in the quote above). Speaker A then summarizes B's account by saying that this was their special time. The preposition *during* in the right-dislocated element is not really necessary and the noun phrase *meal time* or *meal times* would do equally well.

The dislocated elements categorized as pronouns can be distinguished into two groups, one comprises demonstrative pronouns and the other personal pronouns. As for the variant with a demonstrative pronoun, we typically find final *that*, sometimes also *this* and *those*. Most frequently they occur by themselves, rarely also with an operator (expanded RD). Example sentences are given in (4.43).

(4.43) a. No not yet <,> it's very hot **that** (ICE-IND:S1A-020)
b. When was it made **this**? (ICE-GB:S1A-019)
c. They're all really great **those** (ICE-CAN:S1A-069)
d. It's funny **that is** <,,> (ICE-GB:S1A-041)

In the present study, such RD tokens are particularly common in the speech of British English and Indian English speakers, accounting for 20.0% and 17.2% of all RD tokens,

respectively. In the other varieties the proportions are much smaller. Jamaican English shows the next highest proportion with 10.5% of dislocated demonstrative pronouns; in all the other varieties they make up less than 9%.

The dislocated demonstrative pronouns in these RD tokens probably often serve a deictic function. In sentence (a), for example, the interlocutors have a cup of tea while recording their conversation. One could very well imagine the speaker pointing at the cup in front of her while uttering the sentence *it's very hot that*. In addition, the final demonstrative pronoun may also have an emphatic connotation, stressing in (a), for example, that it is the tea that is hot.

While most of the right-dislocated demonstrative pronouns occur on their own, the variant with a personal pronoun normally co-occurs with an operator. These expanded forms of RD will be discussed in some more detail in the following section.

4.2.5 Expanded RDs

Examples of expanded RDs are provided in (4.44). These expanded forms of RD typically serve another function than the RD proper. While RD tokens usually have a clarifying or emotive/evaluating function, as has been pointed out above, the dislocated elements in expanded RDs (or tags, as I will also call them) mainly serve to put stress on the proposition of the clause.[28] In this respect note the co-occurring expressions *sure* and *really* in examples (b) and (c), which strengthen the emphasis that is put on the proposition by the tags.

(4.44) a. They're so freaky **they're** (ICE-PHI:S1A-016)

b. You're not interested in anything but music **sure you're not** (ICE-IRE:S1A-016)

c. It tastes good <,> **it really does** (ICE-PHI:S1A-080)

d. You're very careless with your wallet **so you are** (ICE-IND:S1A-003)

As can be seen in these examples, expanded RDs are also often associated with some form of evaluation, just like 'canonical' RDs. In the sentences in (4.44) this is expressed by the adjectives *freaky*, *not interested*, *good* and *careless*. In contrast to (most) 'canonical' RDs, however, the dislocated elements additionally serve to put emphasis on the proposition. In a sense the evaluation gains in strength through the tag. Timmis (2009) argues in a similar way. Commenting on previous research, he writes that the emphatic potential of right-dislocated elements seems "to be very closely related to the emotionally coloured and evaluative aspect of tails rather than to constitute a separate function" (2009: 337). In my opinion, this seems to be the case particularly with expanded RDs but can, of course, also apply to 'canonical' RDs. What these examples

28 It is debatable whether the expanded form of RD should be discussed in a study on information-packaging constructions at all since such tokens do not primarily serve an information-structuring function. Following previous research (cf. Durham 2011), it was nevertheless decided to include them in the present study. For one, expanded RDs are a type of RD, although not serving the prototypical function. And second, the investigation yields interesting results from a cross-varietal perspective.

teach us is that the clarifying, emotive and emphatic functions are not mutually exclusive but that one RD token can serve several functions simultaneously, with one or the other function possibly predominating in certain contexts, however.

Note that the dislocated element in sentence (d) contains the particle *so* along with the noun phrase and operator. This specific type of expanded RD, which I will call *so*-tag, is particularly common in Irish English and will be discussed in some more detail below.

Table 4.9 gives the frequencies of expanded RD tokens and their percentages out of all RD tokens in the nine varieties of English analyzed. The highest number of expanded RDs can be attested for Irish English, where they account for more than half of all RD tokens (51.6%). The great majority of these are made up of *so*-tags (45 tokens or 70.3%), which seem to be a characteristic feature of Irish English indeed because in the other varieties *so*-tags are very rare. In fact, the only varieties of English for which *so*-tags can also be attested are New Zealand English (2 tokens), British English and Singapore English (1 token each).

Table 4.9: Expanded RDs (absolute token frequencies and percentages out of all RD tokens).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
expanded RDs	11	64	7	-	2	4	2	6	-
%	12.2	51.6	8.4	-	4.2	13.3	10.5	10.3	-

While there are no expanded RD tokens in Canadian and Hong Kong English and only small proportions in New Zealand and Singapore English, they make up more than 10% of all RD tokens in all the other varieties analyzed. Note, however, that with the exception of Irish and perhaps British English the token frequencies are so small that the percentages must be considered with caution. They are possibly not representative of the respective variety of English and should be substantiated by further research based on larger datasets.

Expanded RDs are always formed around a subject rather than an object or complement. This is also the reason why Durham (2011), in her study of RD constructions in a corpus of interviews collected in the city of York in 1996, concentrates on subject RDs only. In order to compare the findings of the present study with Durham's results the following investigation is also restricted to subject RDs. This is meant to shed some more light on the distribution and usage of the expanded form of RD in comparison to 'canonical' RDs.

Following Durham (2011), the following subsections will examine the form of the subject in the dislocation and the type of verb in the main clause. Furthermore, the distribution of RD tokens according to age and gender of the speakers will be examined. Part of the investigation will be restricted to the Irish English data because, as has been seen above, it is the only variety for which a substantial number of expanded RD tokens can be attested.

Subject type

In all nine varieties, right-dislocated nouns occur almost exclusively in 'canonical' RDs. Some rare instances of nouns can be found in expanded RDs though: 9 tokens out of the 96 expanded RDs that have been found in the nine varieties of English analyzed. Examples are given in (4.45).

(4.45) a. He's got some goodies yeah actually um <,> **John has** yeah (ICE-IRE:S1A-013)

b. He was a real actor you know **Old Vic was** (ICE-PHI:S1A-088)

c. I think it's it'll be much easier you know **Chinese is** like (ICE-SIN:S1A-046)

As can be seen in these examples, dislocated nouns in expanded RDs can be proper names or common nouns. Like nouns, dislocated demonstrative pronouns are also almost exclusively found in 'canonical' RDs. Only four instances of demonstratives in the expanded form of RD can be attested, all of them in the speech of British English speakers (cf. e.g. sentence (4.43d) above). Personal pronouns, on the other hand, occur almost always in expanded RDs. The six instances that are found in 'canonical' RDs are of a different kind, as the sentences in (4.46) illustrate.

(4.46) a. No I'm very professional **me** (ICE-IRE:S1A-024)

b. Ours also same thing because it comes in the same line **yours and mine** (ICE-IND:S1A-053)

c. Would they fit me **any of them** <,,> (ICE-IRE:S1A-099)

d. She was telling me next week probably we could meet up **me and her** (ICE-SIN:S1A-094)

Note that the personal pronouns in these 'canonical' RDs differ from the pronouns in expanded RDs. We have oblique forms in (a), (c) and (d), with that in (c) furthermore modified by a quantifier, and two conjoined possessives in (b). The personal pronouns in expanded RDs, on the other hand, are always in the nominative case, which is in fact expected if a pronoun is co-referential to the subject of the clause.

The tendency of personal pronouns to occur in expanded RDs has also been found by Durham (2011: 274).[29] She gives three possible explanations for this observation. First, pronominal dislocated subjects behave differently from nominal ones because they tend not to provide any additional information. Second, Durham finds that the pronouns that occur in 'canonical' (and reverse) RDs are of another kind than those in expanded RDs. That is, while all pronouns in expanded RDs are in the nominative case in her data, those in 'canonical' and reverse RDs are in the oblique form or of a coordinated type (*Mum and I* or *me and Graham*). These findings are very much in line with those of the present study, as has been noted above. And third, if the dislocated subject is a pronoun the reverse RD "could be confused with some types of tag

29 In addition to 'canonical' and expanded RDs, Durham considers a third type of RD in her study. In these 'reverse RDs', as she calls them, operator and noun phrase are inverted: *She got a great bargain **did her Mum*** (Durham 2011: 261). This type of RD was not examined in the present study as there were no such tokens found in the data.

questions" (Durham 2011: 275). I would add a fourth point, which is connected to the first argument, however. As has been pointed out above, expanded RDs tend to serve another major function than 'canonical' RDs. Since pronouns give no additional information about the referent such RD tokens cannot serve a clarifying function. Rather, the repetition of the subject pronoun serves to put emphasis on the subject. In order to put emphasis on the whole clause or its proposition the operator is also repeated. And this is precisely the major function of expanded RDs.

Verb type

Another feature that was investigated in the present study concerns the type of verb in the main clause of 'canonical' and expanded RDs, including auxiliary and modal verbs, as in (4.47a), *be*, as in (4.47b), *have*, as in (4.47c), and full verbs, as in (4.47d).

(4.47) a. I **can't** wait to do it <,> so I can't you know (ICE-IRE:S1A-072)

b. He'**s** absolutely charming Will (ICE-GB:S1A-027)

c. Jesus I'**ve** no messages at all now so I haven't (ICE-IRE:S1A-067)

d. Most of it **comes** from Texas the gas (ICE-GB:S1A-088)

Previous studies have found that the verb *be* used as a main verb is the predominant variant. It accounts for 66% of the RD tokens in Melchers' (1983) study, while full verbs make up 22% and modal and auxiliary verbs 12% (quoted in Durham (2011: 275)). Durham (2011) finds the same overall preponderance of the verb *be* in her data (66%). Modals account for 4% of her tokens, *have* for 9% and other verbs for 21%. The distribution of RD tokens according to verb type in the present data is represented in Table 4.10.

Table 4.10: Distribution of RDs by verb type (token frequencies and percentages of all RD tokens).

	modal		**be**		**have**		**full**	
	N	%	N	%	N	%	N	%
GB	8	9.9	58	71.6	2	2.5	13	16.0
IRE	24	21.6	70	63.1	2	1.8	15	13.5
NZ	13	18.6	45	64.3	2	2.9	10	14.3
CAN	7	25.0	15	53.6	3	10.7	3	10.7
SIN	3	10.0	22	73.3	1	3.3	4	13.3
PHI	1	4.3	16	69.6	-	-	6	26.1
JAM	2	12.5	13	81.3	-	-	1	6.3
IND	7	14.9	35	74.5	2	4.3	3	6.4
HK	1	5.3	13	68.4	-	-	5	26.3

As can be seen, there is some variation across the varieties analyzed. The verb *be* occurs most frequently in RD tokens in all nine varieties, but while it accounts for

more than 81% of tokens in Jamaican English, it makes up only 53.6% in Canadian English. The proportion is so small in the latter variety because auxiliary verbs and *have* occur relatively more frequently. Hong Kong English and Philippine English speakers stand out among the varieties of English analyzed in that they use full verbs more frequently (more than 26% in both varieties). It is difficult to draw conclusions from these results, however, because we are here dealing with very small token frequencies, ranging from only 3 to 7 tokens. What can be noted in comparison to Melchers (1983) and Durham's (2011) results, however, is a general tendency of the ICE speakers to use less full verbs and slightly more modal verbs in RD tokens.

In addition to the overall distribution of RDs by verb type, Durham (2011) also compares the preferred variants across the different forms of RD tokens (in her study, 'canonical' RDs, expanded RDs and reverse RDs). She finds that the reverse RD form is the preferred variant for all verb types but modal verbs, for which the expanded form of RD is the preferred variant. With all other verbs the expanded form is the variant which is used least frequently, that is, all verb types but auxiliaries also occur more frequently in 'canonical' RDs. In the present study, a comparison of verb types across RD forms makes sense only for Irish English because the frequencies of expanded RDs are too small in the other varieties to get conclusive results. Looking at the 111 subject RDs in Irish English that contain a verb in the main clause, it turns out that the expanded form of RD is the preferred variant with full verbs (10 tokens of expanded RDs vs. 5 tokens of 'canonical' RDs) and especially modal and auxiliary verbs (17 vs. 7 tokens). The verb *be* as a main verb, on the other hand, selects both forms of RD to roughly the same extent (34 expanded RD tokens vs. 36 'canonical' RD tokens). The verb *have* is rarely used in general (1 token each). The tendency of auxiliary verbs to occur in the expanded form of RD is in line with Durham's (2011: 275) findings. In contrast to the present study, however, full verbs occur more frequently in 'canonical' RDs in her data.

Durham tentatively suggests that the tendency of auxiliary verbs to occur in expanded RDs may be due to the fact that in many of these sentences the subject is a pronoun, which, as was noted before, are more likely to occur with expanded RDs. The results of the present study support this assumption because all sixteen expanded RD tokens with an auxiliary have a pronominal subject; among the seven 'canonical' RDs with an auxiliary only one has a pronominal subject. It might be interesting to analyze whether auxiliary verbs generally tend to co-occur with a pronominal subject rather than a nominal or clausal one. But this is beyond the scope of the present study.

Gender

In addition to the intra-linguistic features of subject type and verb type, Durham (2011) also examines the distribution of RD forms according to two extra-linguistic features, namely age and gender. She finds that men in York use slightly more RD tokens than women do, but the difference is not statistically significant. Men and women also do not significantly differ in their selection of RD forms, with reverse RDs being the preferred variant for both (57% and 62% of RD tokens, respectively),

followed by 'canonical' RDs (33% and 27%). Expanded RDs are the least favoured variant, making up only 12% of RD tokens for men and 15% for women (Durham 2011: 267).

Because of the low numbers of tokens of expanded RDs in most varieties of English analyzed, the present investigation will again be restricted to the Irish English data. What complicates comparisons across age and gender in ICE-Ireland, however, is the fact that the corpus is not balanced in this respect. That is, younger speakers and female speakers are overrepresented in the direct conversation files, as can be seen in Table 4.11. While there are 127 female speakers aged 19 to 25, the corpus contains only 2 male speakers aged 42 to 49, for example.

Table 4.11: Distribution of speakers in the direct-conversation files in ICE-Ireland by age and gender.

	19-25	26-33	34-41	42-49	50+	nag*	total
women	127	61	9	7	43	34	**281**
men	36	8	6	2	14	20	**86**
total	**163**	**69**	**15**	**9**	**57**	**54**	367

*nag = no answer given

Given this imbalance in ICE speaker proportions, the following results will always be reported in relation to these corpus ratios because a larger number of speakers probably also produces a larger number of tokens. Hence, Figure 4.8 plots not only the proportions of 'canonical' and expanded RDs according to gender but also the ICE speaker proportions of men and women.[30]

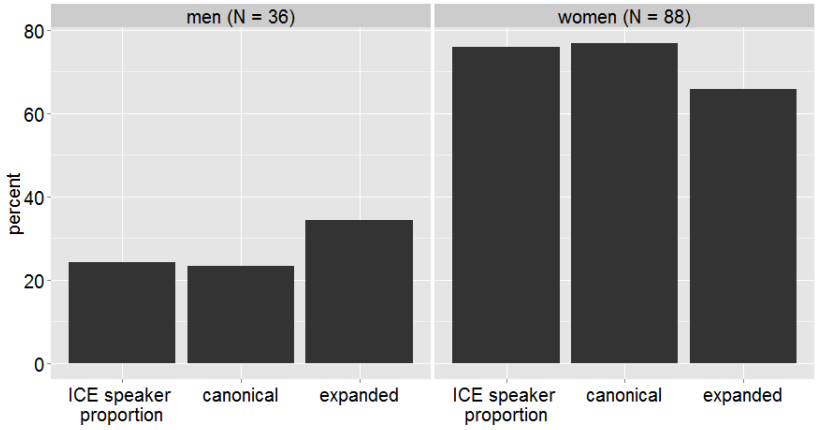

Figure 4.8: Distribution of RD forms by gender in the S1A-files of ICE-Ireland (percentages out of all RD tokens).

30 The corresponding token frequencies and percentages are given in Appendix 6.7.2.

As can be seen, men and women use 'canonical' RDs at similar rates. This can be seen by the fact that the proportions of 'canonical' RDs and ICE speakers are roughly the same: women represent 76% of the data and 77% of the 'canonical' RD tokens come from them, while men represent 24% of the data and use 23% of the tokens. With expanded RD tokens, on the other hand, we find an overrepresentation in the speech of men, with 34% of the tokens coming from them.

In sum, the findings of the present study differ from Durham's (2011: 268) results in that we have more expanded RDs in the speech of men in the present data while in Durham's data it is the women who use the expanded form slightly more often. This might be due to an 'outlier' in the present data. There is one young man who is responsible for nearly one third of all *so*-tags (these will be discussed in some more detail below). When the tokens of this 'high hitter' are removed from the data, the vast majority of expanded RDs comes from women (84%).

Age

The distribution of the two RD forms by age is represented in Figure 4.9, along with the ICE speaker proportions of the five age groups.[31] The probably most striking observation is that both 'canonical' and expanded RDs are overrepresented in the oldest age group (50+). While speakers of this age group represent only 18% of the data, they contribute 24% of all expanded RDs and even 30% of all 'canonical' RDs.

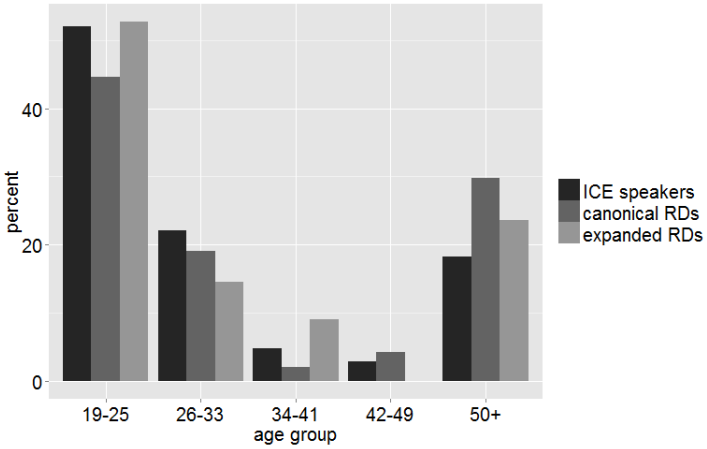

Figure 4.9: Distribution of RD forms by age in the S1A-files of ICE-Ireland (percentages out of all RD tokens).

The somewhat higher proportion of expanded RDs in the youngest age group (19-25)

31 The calculations include only those tokens for which the age is given. Unfortunately, quite a number of speakers did not give the age group they belonged to (cf. category 'nag' in Table 4.11) and therefore, as far as the distribution of RD forms by age is concerned, no conclusions can be drawn from the tokens they contribute. The token frequencies and percentages of RD tokens according to age group are given in Appendix 6.7.2.

is probably again due to the 'high hitter', the young man who uses so many *so*-tags in his speech, as mentioned above. When his tokens are removed from the data, the oldest age group also contributes the largest proportion of expanded RDs (32% of all expanded RD tokens). Somewhat surprisingly, expanded RDs are also overrepresented in the speech of middle-aged people (34-41): while they represent only 5% of the data, they contribute 9% of all expanded RD tokens.

In sum, the findings of the present study tie in with Durham's (2011) results as far as the overall distribution of RD tokens is concerned. She also finds that the oldest age group uses RD tokens most frequently (Note, however, that her oldest age group comprises speakers aged 70 and older.). With respect to the different RD forms, the three age groups in Durham's data show no significant difference in distribution.

So-tags in Irish English

It has been noted above that *so*-tags seem to be a characteristic feature of Irish English since hardly any instances of this type of RD can be attested for the other varieties of English analyzed. Comments on the construction in the literature are sparse. The only note I found is by Harris (1993) who writes that "emphatic sentence tags" are a further focusing device that are worth discussing in the context of Irish English. Unfortunately, he does not analyze them in detail but only notes that "a detailed discussion of these and other discourse features would take us beyond the scope of this short chapter" (1993: 176).

The construction typically has a pronominal subject in the dislocated constituent, some rare instances with pleonastic *there* can also be attested, as illustrated in (4.48).

(4.48) a. He's a real pet **so he is** (ICE-IRE:S1A-049)

b. There's going to be a few sitting and more standing **so there is** (ICE-GB:S1A-072)

Some more variation can be found in terms of the verb in the dislocated element. Usually, the verb from the main clause is repeated, as in (4.48). Hence, if there is a modal or auxiliary verb in the main clause it also appears in the dislocated element, as in (4.49a) to (4.49c); if there is a full verb, we find a form of *do* in the tag, as in (4.49d); if there is negation in the main clause, it is also repeated in the tag, as illustrated in (4.49a).

(4.49) a. I **can't** wait to do it <,> **so I can't** you know (ICE-IRE:S1A-072)

b. I **used to** have great craic with them in England <,> **so I used to** (ICE-GB:S1A-087)

c. Oh he**'s** driving a bus now **so he is** (ICE-IRE:S1A-072)

d. Oh aye I **got** a pass **so I did** (ICE-GB:S1A-094)

What is interesting to note about these structures is not only that they occur much more frequently in Irish English than in the other varieties analyzed but also that they

account for more than half of all RD tokens in this variety (70.3%). The preponderance of *so*-tags may well be due to influence from Irish. The Irish language has the expression *cinnte* 'certainly', which is also often found clause-finally and used for emphatic purposes. For example, the Irish sentence *Beidh mé ann, cinnte* can be translated into the English sentence *I'll be there, certainly* or *I'll be there, so I will.*

Addressing the question of who precisely uses *so*-tags in Irish English, the age and gender profiles of the speakers have been considered. Furthermore, the frequency of use in the Republic of Ireland and Northern Ireland has been compared. It turns out that there is no great difference in distribution in this respect, with speakers from the Republic contributing 25 *so*-tags and speakers from Nothern Ireland 20 tokens. As for the distribution of *so*-tags according to age and gender, the token frequencies are given in Table 4.12. Keeping in mind that young women are largely overrepresented in the direct conversation files of ICE-Ireland, it can be noted that it is predominantly younger men and older women that use *so*-tags.

Table 4.12: Distribution of *so*-tags in Irish English according to age and gender.

	19-25	**26-33**	**34-41**	**42-49**	**50+**	**nag***
women	5	7	2	-	9	6
men	14	-	2	-	-	-

*nag = no answer given

It is important to remember that there is one young man who contributes almost all of the *so*-tags used by men, as has repeatedly been pointed out throughout this subsection. The fact that this young man accounts for 14 tokens of the overall 45 *so*-tags, of course, weakens the representativeness of the findings. This young man comes from County Mayo in western Ireland; he may even have grown up in or have connections with the Gaeltacht, the Irish-speaking region in the county. He may be aware of the fact that he is being recorded and wants to sound particularly Irish. If this was the case this would, on the other hand, speak in favour of the hypothesis that the *so*-tag is a characteristic feature of Irish English. The small frequency of tokens unfortunately does not allow for more reliable conclusions to be drawn, but it would definitely be interesting to follow up the question of whether *so*-tags are a characteristic feature of Irish English that is only used by elderly speakers - and thus constitutes a rather old-fashioned or local feature - or whether it is seen as a marker of Irish identity and is therefore also (again?) used by younger speakers.

The pattern whereby traditionally local features are regaining ground in the youngest generation - and thereby lead to a v-shaped age pattern - is also reported in previous studies. Dubois and Horvath (1999), for example, find such a change in Cajun English, with men increasingly using Cajun variants that had formerly been stigmatized and stereotyped. Women seem not to take part in this change, which Dubois and Horvath explain primarily by the fact that young women "have fewer reasons than do young men to associate themselves linguistically with the current understanding of Cajun identity, which is largely masculine" (1999: 307).

Durham (2011) argues in a similar way. She also finds a v-shaped age pattern for

reverse RDs in the speech of the male speakers in her data. That is, it is especially the older and the young men who use reverse RDs frequently while the middle age group uses them far less often. Since reverse RDs are considered as a primarily northern British feature, Durham suggests that "its high use by the young men of York is another case of a Northern feature being used to signal local identity" (2011: 273).

The *so*-tag may also be such a feature that signals local or Irish identity. The findings of the present study point in this direction, yet they are too sparse to argue conclusively.

4.2.6 Summary

The major aims of this section on the RD construction were to describe its nature and distribution across the nine varieties of English analyzed and to find possible explanations for the observed usage patterns. As for Irish English, explaining the high frequency of RD tokens is quite straightforward: expanded RDs and in particular *so*-tags are used much more frequently by Irish English speakers than by the speakers of the other varieties analyzed.

The relatively high frequency of RD tokens in British English is more difficult to account for. The analyses of different linguistic features - the distribution of tokens that serve an evaluating function, the syntactic function of the dislocated constituent, its realization, the verb type and the distribution of expanded RDs - show that there is no single feature or type of RD that could explain the high number of RD tokens in comparison to all other varieties but Irish English. It has been observed that we mainly deal with 'default' RD tokens, that is, subject RDs with right-dislocated noun phrases serving an emotive or evaluating function. Additionally, we find a relatively high number of right-dislocated pronouns, in particular demonstrative pronoun *that*. Furthermore, British English speakers use more dislocated clauses than most other speakers. Note that for any of these features but final *that* does British English show the highest proportion among the English varieties analyzed. However, it is always among those varieties that are at the top. Taking all these features together, then, we can say that there is a robust number of prototypical RD tokens, but the construction is also quite commonly used with less typical forms and functions. Hence, it is probably the combination of these features that are responsible for the high token frequency in British English.

The surprisingly low frequency in Canadian English is also difficult to explain, as is often the case with something that is just not there. The data suggest that for Canadian English speakers the emotive or evaluating function does not play such an important role, especially in comparison to the other L1 English speakers. Consequently, there is less motivation to use the RD construction.

Indian English and Singapore English show somewhat higher frequencies of RD tokens than the other L2 varieties. For the latter variety this is very likely due to the high number of tokens in which the co-referential pronoun is omitted. In Indian English, we find a relatively high number of right-dislocated demonstratives. These constructions serve a deictic or emphatic function rather than an emotive or clarify-

ing function. This suggests that Indian English speakers use RD tokens more flexibly and for more purposes in discourse than do the other L2 English speakers.

Coming back to Irish English, the close investigation of expanded RDs reveals that it is particularly elderly speakers that use the construction. With respect to *so*-tags, it can be observed that the structure also frequently occurs in the speech of young speakers. Since most of these tokens are contributed by a single young men, this finding is not very robust and maybe not representative of the speech community. Further research is needed to test whether we are here dealing with a situation where a traditionally local and maybe stigmatized feature is gaining momentum in the speech of young speakers. Such an explanation appears quite plausible because similar changes have been reported by other studies (e.g. Dubois/Horvath 1999; Durham 2011), and in these studies it is also the young men who are in the lead of such changes. However, the observed pattern in the present study may, of course, just as well be due to an idiosyncratic feature of the speech of this young man.

4.3 Fronting constructions

This section examines fronting constructions, that is, constructions in which an element that would normally occur post-verbally is placed in initial position. A speaker may, for example, use the sentence *The Lord's Prayer I've translated* instead of *I've translated the Lord's Prayer* for some pragmatic reason. As noted earlier (cf. section 4), fronting constructions occur much more frequently in the speech of Indian English speakers than in that of the speakers of the other varieties. Hence, one of the major questions addressed in this section is what motivates Indian English speakers to use fronting constructions so much more frequently than the other speakers do.

The structure of the section is as follows. First, the overall distribution of fronting constructions across the nine varieties of English analyzed will be presented, followed by the investigation of a number of syntactic and pragmatic features.

4.3.1 Overall distribution

Figure 4.10 shows the normalized frequencies per 100,000 words of fronting constructions in the nine ICE samples analyzed. As can be seen, fronting is a rather rare phenomenon in English (except for Indian English), a finding which has already been reported in previous research (cf. e.g. Netz/Kuzar 2007, 2011). Netz and Kuzar (2011) argue that it is "[b]ecause of the relatively rigid word order of English, and the fact that in English the 'formal sentence structure' overrides the 'functional sentence perspective'" that "in English the OF [object fronting] construction exhibits an extremely limited distribution" (2011: 155). While Indian English shows a very high frequency of fronting constructions, in most of the other varieties the number of tokens is in fact

so small that reliable conclusions can often not be drawn, as the investigations below will show. In detail, the frequencies per 100,000 words are as follows: ICE-India 98.6; ICE-Singapore 26.1; ICE-Ireland 25.3; ICE-Great Britain 17.4; ICE-New Zealand 15.7; ICE-Canada 13.3; ICE-Philippines 11.5; ICE-Jamaica 11.3; ICE-Hong Kong 5.9. The differences between Indian English and the other varieties are statistically highly significant ($p < 0.001$), according to chi-squared tests.

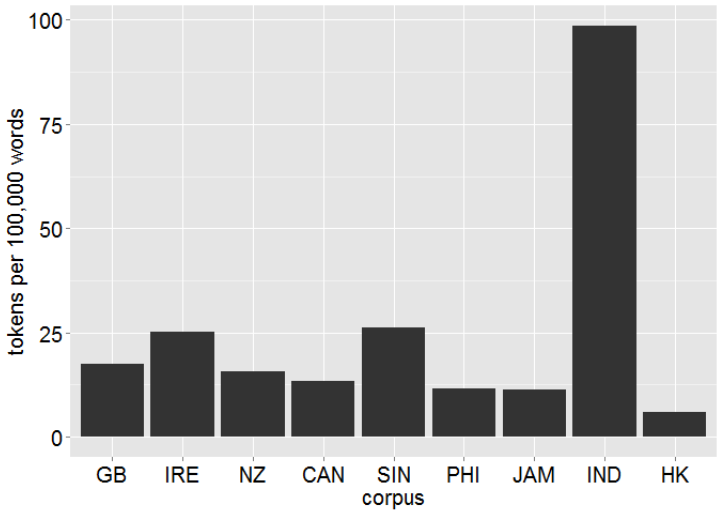

Figure 4.10: Fronting constructions in nine ICE corpora (frequencies per 100,000 words).

These findings tie in well with a number of previous studies. Gregory and Michaelis (2001), culling their data from the syntactically parsed portion of the Switchboard Telephone Speech Corpus (American English), find around 17.6 fronting tokens per 100,000 words. This figure is very much in line with the findings in the British component of ICE (17.4 tokens). As for Irish English, the somewhat higher frequency of fronting constructions in comparison to the other L1 English varieties does not come as a surprise. Although IT-clefts are commonly reported as the major focussing or topicalization device in Irish English, fronting constructions are also described as a characteristic feature of the Celtic Englishes (especially Welsh English; cf. e.g. Filppula 1999, 2006, 2009; Hickey 2007). The findings of the present study furthermore support claims stating that the construction is a common feature of Indian English (Bhatt 2004, 2008; Lange 2012; Sharma 2012b).

In the following paragraphs, a number of syntactic and pragmatic features will be examined in order to find possible explanations and motivating factors for the high frequency of fronting constructions in Indian English and to reveal qualitative and/or quantitative similarities and differences among the English varieties analyzed.

4.3.2 Topic continuity

Examining fronting constructions in Indian English, Lange (2012: 134) argues that there are many examples that "display an explicit discourse-linking function". This discourse function is highly reminiscent of the function she identified for left dislocation constructions in Indian English. What is different in the present context, however, is that the group of items that have an explicit discourse-linking function contains not only preclausal elements that take up an element from the immediately preceding discourse - as was the case with LD constructions - but also items whose fronted element is a demonstrative pronoun or a noun phrase preceded by a demonstrative. Examples of the first type are given in (4.50) and (4.51). The two other types are exemplified in (4.52) and (4.53) below.

(4.50) B: The thing is that it is rural area only <,,> but uh <,> the people <,> are from all over India <,> uhm <,> they are staying here <,,>

A: Uh <,> getting **cosmopolitan**

B: **Cosmopolitan** it is <,,> it is bit like that <,,> So <,> in this way our colony <,> uh <,> Aditya Nagar is <,,> Aditya Nagar is just like a <,,> mini-India <,> [laughs] I can say

(ICE-IND:S1A-063)

(4.51) B: So you are still continuing with **Wagner** you haven't thought <,>

A: No **Wagner** I use for the advanced diploma course <,> third year course <,> first year second year I follow this *Russian for everybody* <,> that's what I did do with my students now <,> same batch which I now <,> and then uh I have switched over to Wagner now <,> now they are doing on Wagner <,> in higher diploma

(ICE-IND:S1A-060)

In these examples, the major motivation for preposing seems to be the creation of topic continuity. In (4.50), speaker B picks up the word *cosmopolitan* from the immediately preceding utterance and places it in initial position to draw a link to the previous discourse and to show agreement with speaker A. Furthermore, politeness seems to play a role because many of the examples are question-answer pairs, as in (4.51), with the answer repeating the topic noun phrase of the question.

The types of fronting involving a demonstrative pronoun in the fronted element are illustrated in (4.52) and (4.53).

(4.52) C: If she has a passport it's well and good otherwise great problem *yaar*

A: **That** she has I think

(ICE-IND:S1A-037)

(4.53) A: We need to continue *na*

B: Yeah <,> we must continue other ten minutes [laughter]

This exercise we have to be pulled on [cough]

(ICE-IND:S1A-045)

These examples illustrate that the preposing of demonstrative pronouns or of noun phrases with a demonstrative pronoun may also serve as explicit links in the communicative situation. In (4.53), the noun phrase *this exercise* refers back to the speakers' task of recording another ten minutes of conversation, a task which they seem to find rather difficult or daunting. The preposed demonstrative pronoun in (4.52) does not only refer back to the noun phrase *passport* in the preceding utterance, but probably also receives focal stress. That is, the fronting construction in (4.52) is an example of what is sometimes also referred to as focus preposing. The sentence in (4.53), on the other hand, is an example of fronting proper, with the focal stress probably falling on *pulled*. Lange (2012: 135) argues that it is not the number of preposed demonstrative pronouns as such that is striking for Indian English, but rather that such pronouns occur in focus preposings. Her argument is based on the fact that she found no such items in the British English data she examined. The findings of the present study suggest otherwise. The occurrence of anaphoric pronouns in focus preposings is not particularly remarkable because such items can also be attested for other varieties of English. Consider, for example, the two sentences below, which are from Canadian English and Philippine English, respectively.

(4.54) A: What are you taking next semester?
B: Uhm <,> I have <,,> nutrition
A: Yeah **that** I knew

(ICE-CAN:S1A-098)

(4.55) A: So that's why you hated Bio so much
C: It was difficult *e*
A: Opening of the cockroach
C: Uh **that** I did

(ICE-PHI:S1A-056)

Investigating these three types of fronting construction serving an explicit discourse-linking function, Lange (2012: 137) concludes that "one overarching discourse motivation for topicalization in spoken IndE is to express topic continuity" (she calls fronting constructions 'topicalization'). The findings of the present study show, however, that this motivating factor is not exclusive to Indian English. Rather, it is again Philippine English that sticks out in this respect, as the percentages in Table 4.13 indicate. While 32% of all fronting tokens in the Indian English data have an explicit discourse-linking function, the proportions are similarly high in Singapore English (32%) and even higher in Philippine English (52%) and Hong Kong English (36%).

Given these findings, the creation of topic continuity cannot be seen as the sole motivating factor for Indian English speakers to use fronting constructions. In the present context, the function of creating topic continuity was examined on the basis of the three types of fronting construction outlined above: preclausal elements that

take up the topic of the immediately preceding discourse and preclausal elements that contain a demonstrative pronoun, as the sole element or preceding a noun phrase. In what follows, the notion of 'topic continuity' will be given a wider interpretation.

Table 4.13: Fronting constructions with an explicit discourse-linking function (absolute frequencies and percentages out of all fronting constructions).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
FRON tokens	35	51	36	28	53	25	24	213	14
TC tokens*	9	12	4	7	17	13	5	69	5
TC %	25.7	23.5	11.1	25.0	32.1	52.0	20.8	32.4	35.7

*TC = topic continuity

Like in the case of LD, the preclausal elements will be analyzed according to their information status and their persistence in the following discourse. This investigation is meant to give a better idea of the fronting construction's discourse functions and to compare these functions to those of the left dislocation construction. It has been noted before that the two constructions are structurally quite similar. Both involve a clause constituent that is taken out from its canonical position in the core of the clause and placed in initial position for pragmatic purposes. They differ, however, in that there is a co-referential pronoun in LD constructions. The following discussion will show that they also differ in terms of the information status of the initial element.

Both properties - information status and persistence - again comprise three categories. As for the information status or anaphoricity, there are preposed constituents that pick up elements from the preceding discourse and thus represent discourse-old information. Second, there are items that have not been mentioned before and thus refer to discourse-new information. And finally, there are preposed elements that have implicitly been mentioned in the previous discourse and are therefore inferable from the discourse.

Persistence refers to the question whether and in how far the initial element of a fronting construction persists in the following discourse. In order to rate the persistence, the following five sentences were looked at and the items grouped into preposed elements that are not mentioned in the following five sentences at all and items that are mentioned again in the following turns in form of a pronoun. The third group contains all items that do not belong to any of the other two groups, that is, items that are taken up as a noun phrase in the following discourse or items that are only indirectly talked about afterwards.

Before looking at the distribution of the anaphoricity ratings and persistence scores, it should be pointed out once again that the frequencies of fronting constructions are so small in some corpora analyzed that the percentages plotted in the figures below are not particularly robust. Hence, the results should be taken with more than a pinch of salt and they clearly call for more research based on larger datasets.

Figure 4.11 shows not only the anaphoricity ratings of fronting constructions but also those of LD constructions, plotted here for reasons of comparison (cf. Figure 4.3

in section 4.1).[32] The probably most striking observation that can be made is that fronting constructions refer to discourse-new information far less frequently than do LD constructions (cf. category 'new'). The proportions of initial elements that refer to discourse-old information, on the other hand, are roughly the same for both constructions in all varieties (cf. category 'old'). New Zealand English is an outlier in this respect, but this may simply be due to the low frequency of items. With a few more tokens the distributional pattern might look very different. Inferable information is found more frequently with fronting constructions than with LD constructions (cf. category 'inferable').

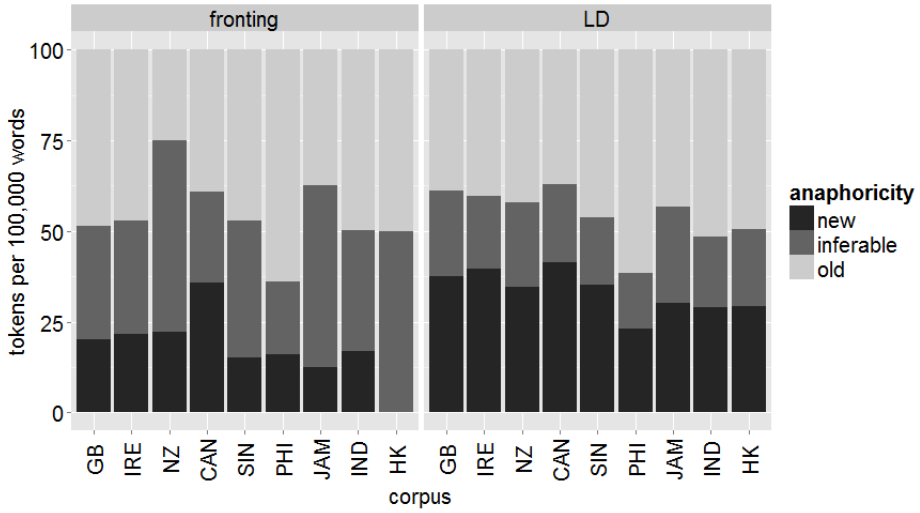

Figure 4.11: Anaphoricity ratings of LD tokens and fronting constructions (percentages of initial elements containing old, new or inferable information).

The finding that fronting constructions overwhelmingly refer to old or inferable information is very much in line with previous studies. Prince (1985), for example, argues that the initial element in fronting constructions (she only looks at noun phrases) is "either already evoked in the discourse or else in a salient set relation to something already evoked in or inferrable from the discourse" (1985: 70). Similarly, Birner and Ward (2004) note that the initial element in fronting constructions "is constrained to be old information. More specifically felicitous preposing [i.e. fronting; cw] in English requires that the information conveyed by the preposed constituent constitute a discourse-old anaphoric link to the preceding discourse" (2004: 159). Note that Ward and Birner's group of items that create discourse-old anaphoric links also includes the fronting constructions that are here rated as 'inferable'. Furthermore, while Ward and Birner claim that the preposed element always refers to old information, the findings of the present study indicate that this is not categorically the case, but that the preposed elements may also contain new information.

For spoken American English, Gregory and Michaelis (2001: 1695) also find that

32 The token frequencies and percentages of the anaphoricity ratings of fronting constructions can be found in Appendix 6.8.3; those of LD constructions in Appendix 6.6.5.

fronting constructions contain new information much less frequently than LD tokens do; the difference in their data is in fact even more striking, with 62% of LD constructions referring to new information but only 25% of fronting constructions.

A further point worth mentioning concerns the category 'inferable'. As noted earlier, the proportions of this category are much higher in the case of fronting constructions than in that of LD tokens in all nine varieties of English analyzed. This is due to the fact that many fronting constructions involve partially ordered set or poset relations, for example, whole/part or type/subtype relations (cf. Prince 1998; Birner/Ward 1998). The following examples are given for illustration. In (4.56), the speakers talk about different Christian, Muslim and Hindi festivals in India.

(4.56) A: And then their festivals like *Ganesh* <,>
B: *Chaturthi* and
A: *Chaturthi* yeah and then <,,> uh *Diwali*
B: *Diwali* yeah ***Holi*** do you have
A: Yeah ***Holi*** also we have

(ICE-IND:S1A-065)

In this example, there is a type/subtype relation. The speakers enumerate a number of festivals that come to their minds, so 'festivals' is the superordinate term and *Holi*, fronted in speaker B's second turn, is a subtype of the type 'festivals'. Through this poset relation the term *Holi*, while mentioned for the first time in B's second turn, is linked to the preceding discourse and thus topic continuity is created.

The following example illustrates a set/subset relation. The set comprises all the people that Matt, a common friend of the interlocutors, does not read very well and among these he has particular problems with girls, the subset.

(4.57) B: Matt doesn't read people <,> male or female <,> very well
A: Well neither do I
B: He doesn't <#> But I mean he he ju <,> he <,> **especially girls** he doesn't read very well <#> Matt's naive in certain ways you know

(ICE-IRE:S1A-018)[33]

In addition to fronting constructions that involve such poset relations, the category 'inferable' also includes a number of items that summarize, evaluate or comment on what has been said before.

(4.58) T: Glen's film showed before *Bad Lieutenant*
And he had to get up and make a speech in front of
X: Did he
T: Yeah in front of this packed house
I would like to thank blah de blah de blah

33 The mark-up <#> indicates a speech unit.

X: [clears throat] I thought he didn't turn up to it

A: He did but he didn't tell anyone that it was on or that he was showing before it or anything

X: Oh I see

A: **Sneaky one** he is

(ICE-NZ:S1A-026)

The speakers in this piece of discourse talk about Glen, a common friend or at least a person they all know. Speakers T and A tell X that Glen turned up at some event not telling anyone that he would do so. The topic Glen is concluded by A's utterance *Sneaky one he is*, summarizing or commenting on Glen's behaviour.

The discussion so far has concentrated on the differences between fronting and LD constructions, revealing common tendencies among the nine varieties of English analyzed: the preposed elements of fronting constructions refer less frequently to discourse-new information than do those of LD constructions, but they have more items with inferable information instead. As far as the differences across the varieties of English are concerned it can be noted that the L1 English varieties show slightly higher proportions of fronting constructions that contain discourse-new information. Recall that a similar finding could also be attested for LD constructions. Furthermore, Philippine English again sticks out by showing the highest proportion of fronting constructions that refer to old information (64% out of all fronting constructions), like in the case of LD constructions. Given these findings, it seems that Philippine English speakers - more so than speakers of other varieties - put elements in initial position by means of fronting constructions and even more so by means of LD constructions to draw links to the preceding discourse and thus create topic continuity.

The somewhat different distributional patterns attested for New Zealand English and Canadian English are rather surprising and for the moment I cannot think of any plausible reason for them. Maybe they are simply due to the overall small frequencies of fronting constructions and one token more or less would lead to a very different picture.

As far as Indian English is concerned, the analysis of the information status of the initial element in fronting constructions does not provide any explanation that could account for the high frequency of use in this variety of English. As can be seen in Figure 4.11, Indian English speakers do not behave differently from the other speakers in any way.

Let us turn to the persistence of the preposed element. For reasons of comparison, Figure 4.12 shows the persistence scores of both fronting and LD constructions (the latter is Figure 4.4 in section 4.1).[34] Comparing the results of these two constructions, very different distributional patterns can again be observed. In fronting constructions, the preposed elements persist in the following discourse less frequently than the initial elements in LD constructions (cf. category 'zero'). Gregory and Michaelis'

34 The token frequencies and percentages of the persistence scores of fronting constructions are given in Appendix 6.8.3; those of LD constructions in Appendix 6.6.5.

(2001: 1696) study on spoken American English yields very similar results: the preposed elements of fronting constructions persist much less frequently than that of LD constructions. These findings reflect one of the major discourse functions of LD constructions, namely topic establishment, with topics usually tending to persist in the following discourse. Fronting constructions, on the other hand, rather serve a discourse-linking function. That is, the preposed element does not establish a topic but rather draws a link to the preceding discourse. Note, furthermore, that it is typically subjects that are preposed in LD but objects in fronting. In English, the subject is often equivalent to the topic and thus it is not surprising that more preposed elements persist after LD constructions than after fronting constructions.

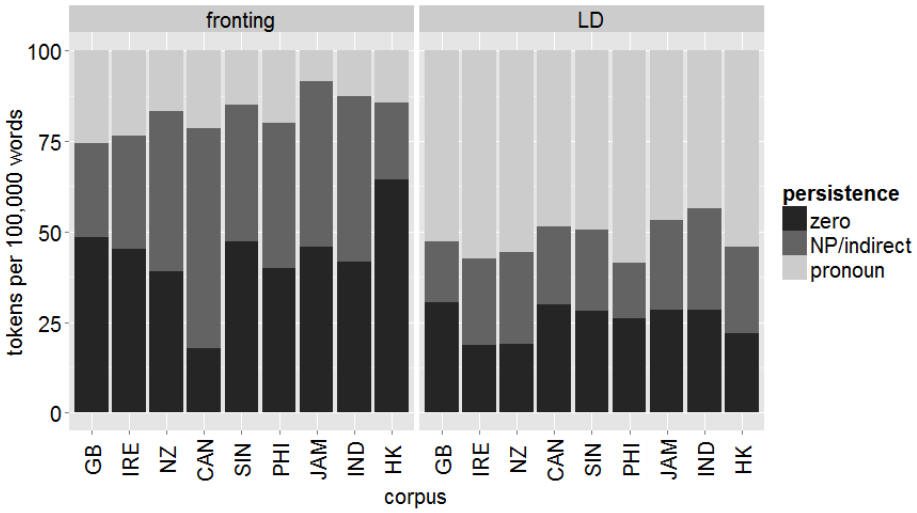

Figure 4.12: Persistence scores of LD tokens and fronting constructions (percentages of initial elements that persist as a pronoun, indirectly/NP or not at all).

Indirectly referred to items can be found more frequently in fronting than in LD constructions (cf. category 'NP/indirect'). This might reflect the fact that fronting constructions are often used in poset relations, as noted earlier. That is, the preclausal element is a member of a set or type which is the theme of the discourse and which continues to be talked about in the ensuing discourse.

Looking at the results for fronting constructions only, it can be noted that L1 English speakers tend to talk about a preposed element in the following discourse slightly more frequently than L2 speakers. Canadian English sticks out by showing a particularly small proportion of elements that are not mentioned at all afterwards. But note that this proportion is based on 5 tokens while that of preposed elements that are pronominally taken up again in the following discourse is based on 6 tokens. These small frequencies call for further research based on larger datasets.

As for Indian English, again no remarkably different behaviour can be noted. That is, the information status of the preposed element and its (non-)persistence into the following discourse cannot be given as explanatory factors for the high incidence of fronting constructions in this variety of English.

4.3.3 Syntactic features

This subsection examines in some detail the syntactic function of the preposed constituent and its realization. This is meant to reveal commonalities and possible idiosyncratic features among the varieties of English analyzed. In particular, it is tested whether Indian English speakers somehow deviate from the other speakers and show more variation, which in turn could explain the high frequency of fronting constructions in this variety of English.

Syntactic function of the preposed constituent

As far as the syntactic function of the fronted element is concerned, we find fronted objects, complements and (obligatory) adverbials in English fronting constructions. Examples of these three types of fronted constituents are given in (4.59).

(4.59) a. **A radio interview** you used (ICE-GB:S1A-047)

b. **Nice** he is (ICE-IND:S1A-093)

c. **Near the coastal area** every <,> yeah <,> time it will come (ICE-IND:S1A-002)

Fronted objects are the most common type in all varieties analyzed, as can be seen in Table 4.14. However, the preponderance is not as pronounced in most varieties as is the case with preclausal subjects in LD and RD constructions (except for Hong Kong English where we have nearly 93% of fronted objects).

Table 4.14: The distribution of fronting constructions according to the syntactic function of the preposed element (absolute token frequencies and percentages out of all fronting constructions).

corpus	fronting	object		complement		adverbial	
	N	N	%	N	%	N	%
GB	35	25	71.4	10	28.6	-	-
IRE	51	23	45.1	17	33.3	11	21.6
NZ	36	20	55.6	12	33.3	4	11.1
CAN	28	17	60.7	1	3.6	10	35.7
SIN	53	45	84.9	5	9.4	3	5.7
PHI	25	20	80.0	4	16.0	1	4.0
JAM	24	17	70.8	4	16.7	3	12.5
IND	213	166	77.9	36	16.9	11	5.2
HK	14	13	92.9	1	7.1	-	-

The proportions of fronted objects are smallest in the L1 varieties Irish English, New Zealand English and Canadian English, where they account for 45%, 56% and 61% of all fronting tokens, respectively. In the case of LD and RD constructions, on

the other hand, the proportions of dislocated subjects are over 70% in all nine corpora. These findings suggest that fronting constructions are generally used more flexibly than LD and RD constructions as far as the syntactic function is concerned. Similar to LD constructions, it is again the L1 English varieties that show most variation.

As for fronted adverbials, it is interesting to note that they account for relatively large proportions in Irish English and Canadian English, making up 21.6% and 35.7% out of all fronting constructions, respectively. Quite a number of these obligatory adverbials involve phrasal verbs in two different ways: either only the particle is put in initial position, as illustrated in (4.60) below, or the whole phrasal verb is placed in initial position plus reversed order of particle and verb, as in (4.61) and (4.62). Admittedly, these sentences are debatable cases of fronting because the fronted elements are rather atypical adverbials, which belong more strongly to the verb than adverbials normally do. But it was decided that they sufficiently resemble cases with proper adverbials to be included in this chapter. Compare, for example, the sentences *She comes home* or *She came to my house* where the adverbials *home* and *to my house* can more easily be identified as adverbials of place. I would argue that *In she comes* and *Home she comes* are very similar in form and function and can therefore here be treated in the same way.

Rather long stretches of discourse are quoted from the Irish and Canadian ICE components in the following extracts because looking at the context it can be noted that this type of construction occurs often in stories, especially in stories which are told in a rather dramatic and rousing or funny way.

(4.60) I was telling the girls at work about that recently <#> This woman she came in and she'd hair up in a bun and a black Granny Moore little dress <#> **In she came** <,> <#> Suddenly it was a stag party <,> she took everything off <#> Now we're talking about wobbly bits [laughter] <#> She had rolls upon rolls <#> She had rolls on her rolls

(ICE-IRE:S1A-082)

Note that the speaker says in her second sentence, using a left dislocation construction, *This woman she came in.* And after a few more words she repeats these words but now with *in* in initial position. Uttered in this way these three words sound more dramatic and entertaining. Also note the following term *suddenly*. This clearly indicates that the speaker wants her story to sound funny and dramatic.

The extract in (4.61) is taken from a conversation between a number of young men. Speaker C tells about a striptease artist who was invited as a special guest to a friend's thirtieth birthday party. Unluckily, the artist arrived while everybody was being engrossed in watching some soccer match and thus showed no particular interest in him. Note that the story again is told in a rather entertaining way.

(4.61) C: And the poor creature arrived in the middle of one of the uhm World Cup matches [...] The poor guy he was kind of pushed off upstairs until the uhm

A: Oh yeah we were watching the Ireland England

C: Until the World Cup finished Ireland <#> And uhm Italy

A: Ireland and Italy <#> Was it <#> No wasn't it Ireland and England

C: No it was the Italy one

A: Was it

C: It was the last one <#> The one they actually lost <#> And everybody was really like <,> well anyway <,> [laughter] <#> **And down comes your man** <#> Nobody was in the humour for him at that stage

(ICE-IRE:S1A-082)

The example in (4.62) is from ICE-Canada. Two speakers, born in 1951, talk about their childhood memories and tell tales about their first encounters with the television. Note that the speakers laugh a lot.

(4.62) A: We had a television and I remember one day there was a woman screaming on the television <,> and I must have been three years old and my mother said that's it

B: [laughs]

A: Cos the three little kids were sitting there watching this

B: [laughs]

C: [laughs] **Out went the television** onto the front lawn with a sign television for sale

B: [laughs]

A: And we never saw it again

(ICE-CAN:S1A-043)

Such examples are not unique to Irish English and Canadian English, but they seem to be more common in these ICE components. In order to find out whether fronted directional and locative particles are a feature of Canadian English and in particular Irish English, as the ICE data suggest, and to find out in which register they are most common, searches in COCA (Corpus of Contemporary American English) and GloWbE (Corpus of Global Web-based English) have been performed. All phrasal verbs with *come* and *go* for which instances with fronted particles could be found in COCA were included in the search.[35] The searches in COCA reveal that the constructions occur by far most frequently in fiction (46%), followed by spoken language (20%), magazines (19%), newspapers (11%) and academic writing (4%). This supports the impression that the construction is mainly used as a stylistic device in stories and in spoken language, as suggested by the ICE data.

Looking at GloWbE, fronted particles in phrasal verbs with *come* and *go* turn out

35 The Online Oxford English Dictionary (OED) was used to get all phrasal verbs with *come* and *go* (cf. "come, v." and "go, v." OED Online; accessed 1 August 2014). After identifying those verbs for which the fronting of the particle is possible, the following searches were performed: "along|back|down|foward|in|off|on|over|up I|you|we|she|he|they|it [come]" and "around|away|back|down|in|on|off|out|over|round|up I|you|we|she|he|they|it [go]". Interestingly, fronted particles are much more frequent with *go* than with *come* (14.1 vs. 1.7 instances per 1 million words).

to be indeed most common in Irish English (6.6 tokens per 1 million words), followed by New Zealand English (6.0), British English (4.8) and Canadian English (4.2). It is surprising to find British English among the varieties with the highest frequencies of fronted particles because no such items could be found in the British ICE component. This, however, may again be due to the higher level of formality of the 'private dialogues' files in the British ICE component as compared to the other corpora.

Turning to fronted complements, it is surprising that there is only one such construction in ICE-Canada while the other L1 English varieties show so much higher token frequencies and percentages. Fronted complements typically occur in constructions with the copular verb *be*, sometimes also with *call*. They are normally realized in the form of adjective phrases or noun phrases to describe a quality of the subject or to identify the subject's possession or name. Through their initial placement the elements are emphasized, as the examples in (4.63) illustrate.

(4.63) a. **Identical** they were. (ICE-IRE:S1A-065)

b. **Oh great** that'd be. (ICE-GB:S1A-042)

c. **Your jeans** were they? (ICE-IRE:S1A-080)

d. **Mark** her husband's called. (ICE-GB:S1A-028)

I can only speculate on why fronted complements occur so infrequently in the speech of Canadian English speakers. It is quite unlikely that Canadian English speakers do not express their feelings and opinions or some quality of an entity, but it is possible that they prefer to use unmarked SVX sentences to do so. Furthermore, they may prefer to use pitch and intonation rather than word order changes to mark certain elements in the clause as important, new or noteworthy. The low frequency may also be due to the make-up of the corpus because when discussing RD constructions it has been observed that Canadian English speakers use RD tokens that serve an emotive or evaluating function far less frequently than the other L1 English speakers. Possibly, the conversations in ICE-Canada do just not give as many opportunities to express feelings and opinions or qualities of some entity.

Realization of the preposed constituent

This section discusses the realization of the fronted element. The variants that are found in the data include fronted noun phrases (nouns and pronouns), prepositional phrases, adjective phrases and clauses. Examples are given in (4.64).

(4.64) a. And **one chapter** I added on the new critics <,,> (ICE-IND:S1A-026)

b. Uh **that** I did. (ICE-PHI:S1A-056)

c. **In Hindi** also we can write <,,> (ICE-IND:S1A-071)

d. **Oh great** that'd be. (ICE-GB:S1A-042)

e. But exactly <,> **exactly how it was slotted in** <,> I can't remember quite (ICE-GB:S1A-012)

The distribution of these variants across the nine varieties of English analyzed can be seen in Figure 4.13.[36] Note that the percentages presented in the figure are not very robust because of very small token frequencies in some of the samples. Consequently, only some major tendencies will be pointed out in the following paragraphs.

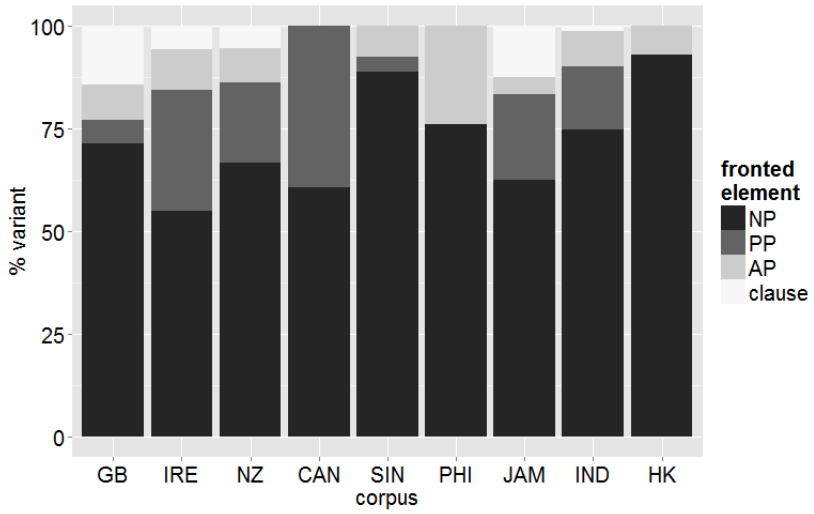

Figure 4.13: The distribution of fronting constructions according to the realization of the fronted element (percentages out of all fronting constructions).

Fronted noun phrases are the most frequent variant, accounting for more than half of all fronted elements in all varieties analyzed. They are particularly dominant in Hong Kong English and Singapore English, where they make up more than 92% and 88%, respectively. The proportion is smallest in Irish English (54.9%). In this variety, along with British English, New Zealand English and Jamaican English, we find the greatest variation with respect to the realization of the fronted element.

Fronted prepositional phrases are most common in Irish English (29%) and Canadian English (39%), which is very likely due to the fact that phrasal verbs with fronted particles are part of this category. Recall that fronted particles in phrasal verbs are quite common in these two varieties of English. Interestingly, fronted prepositional complements, as in *Up on the roof he was when the lightning struck*, are reported in the literature as a particularly common type of fronting in Irish English (Hickey 2007: 267), a claim that can be supported by the present study: prepositional phrases and complements are found relatively frequently in inital position in this variety of English (cf. Figure 4.13 and Table 4.14).

What is furthermore interesting to note with respect to fronted prepositional phrases is their (near-)absence in Hong Kong English, Philippine English and Singapore English (2 tokens or 3.8% in Singapore English; zero in the other two varieties). What may account for the rare incidence of fronted prepositional phrases in these three varieties of English is the fact that prepositions are often omitted in L2 or vernacular va-

36 The exact token frequencies and percentages of the different variants are given in Appendix 6.8.3.

rieties of English (cf. e.g. Platt et al. 1984; Deterding 2007; Mair/Winkle 2012). Since the (potential) fronting constructions with missing prepositions can easily be confused with 'hanging-topic' constructions it was decided to excluded from the present study.[37] To illustrate the difficulty of keeping the two types of construction apart consider the following examples from ICE-Singapore. The conversation is about C's application for a job at IBM.

(4.65) A: Did you apply for the PR job at uh Marina Mandarin
C: No I only applied for
In fact **the IBM one** I applied about a year ago
So when they call me up uh yesterday evening about five
I said **IBM** I don't remember applying
(ICE-SIN:S1A-004)

These two 'hanging-topic' constructions could well be expressed as preposed prepositional phrases in the following way: *For the IBM one I applied about a year ago* and *For IBM I don't remember applying*. Note that speaker C starts out with *applied for* before he interrupts himself and then continues without the preposition. This suggests that the item could probably be classified as a fronting construction rather than a 'hanging-topic' construction. The second example, on the other hand, really seems to be a 'hanging-topic' construction. Although not indicated in the transcription, there seems to be a pause after *IBM* and the expression is less closely linked to the rest of the clause than *the IBM one* in the first example. These examples show that the status of these and similar items is difficult to precisely specify. That is why they have been excluded from the investigation.

Interestingly, it is exactly those three varieties of English whose speakers have a topic-prominent L1 (Mandarin, Cantonese and Filipino/Tagalog) that show such low frequencies of preposed prepositional phrases. Recall that 'hanging-topic' constructions have been described as typical of topic-prominent languages (cf. Lambrecht 2001a and section 3.2.3). Possibly, speakers of these English varieties prefer 'hanging-topic' constructions over fronting. It might be worth analyzing and comparing the two structures more systematically, especially in those varieties of English that come into contact with topic-prominent languages.

A final point that deserves mentioning with respect to the realization of the preposed element concerns the category 'clause'. As can be seen in Figure 4.13, fronted clauses occur rarely or never in the data. What is interesting to note, however, is that along with the three L1 English varieties British English, Irish English and New Zealand English it is only Jamaican English that shows a substantial number of preposed clauses. This observation ties in with previous results in this study. Recall that the analyses of complex initial elements in LD constructions and of complex or extended existential *there*-constructions yield similar results: the L1 English speakers and Jamaican English speakers tend to use more complex constructions more frequently than the speakers of the other varieties.

37 The 'hanging-topic' construction is briefly discussed in section 3.2.3.

4.3.4 Summary

One of the major aims of the present section was to find possible explanations for the high frequency of fronting constructions in Indian English. The function of creating topic continuity, suggested in the literature as a motivating factor, turns out not to be more dominant in the speech of Indian English speakers than in that of other speakers. Similarly, the investigation of the information status of the preposed element and its (non-)persistence in the ensuing discourse yields no results that could conclusively explain the high frequency of fronting constructions in Indian English. Indian English speakers do not behave differently from the other speakers. Rather, it is Philippine English speakers that stand out, using to a great extent fronting constructions whose preposed elements refer to discourse-old information and persist in the ensuing discourse in the form of pronouns. Thus, it seems that the creation of topic continuity is an important motivating factor for Philippine English speakers to use fronting constructions. Recall that the same holds for LD constructions.

The investigation of the preposed constituent's syntactic function and its realization also yields no results that could plausibly account for the high incidence of fronting constructions in Indian English. Rather, we can observe that the speakers of the L1 English varieties show more variation in this respect. As for the realization of the preposed constituent, one particularly interesting finding is the (near-)absence of fronted prepositional phrases in the speech of Singapore English, Hong Kong English and Philippine English speakers, exactly those speakers who have a topic-prominent L1. With prepositions often being omitted in these varieties of English, (potential) fronted prepositional phrases can easily be confused with 'hanging-topic' constructions. Consequently, such ambiguous cases have been excluded from the analysis. Since 'hanging-topic' constructions are described in the literature as characteristic features of topic-prominent languages (Lambrecht 2001a), it might be interesting to examine whether they are indeed more frequent in those varieties of English that are in contact with topic-prominent languages or whether they occur in L2 and vernacular varieties of English generally (prepositions are also often omitted in the latter varieties; cf. e.g. Platt et al. 1984; Deterding 2007).

Coming back to Indian English, I can only agree with Lange (2012) who notes that "the causes or motivations for the higher incidence of topicalizations [i.e. fronting constructions; cw] in spoken IndE compared to spoken BrE, or in New Englishes generally, are notoriously difficult to pin down" (2012: 150). In the present study, no single syntactic or pragmatic feature could be identified that could plausibly explain the large frequency of fronting constructions in Indian English. Furthermore, we have seen that the frequency of the construction is not only overwhelmingly higher in Indian English as compared to British English but also in comparison to the other L2 English varieties. Hence, I disagree with Mesthrie (1992) who rules out substrate influence as impacting on the high incidence of fronting constructions in the New Englishes and opts for "universals of discourse structure" as the decisive influencing factor (1992: 157). On the contrary, I would argue that influence from the substrate languages plays an important role indeed in the use of fronting constructions by Indian English speakers. As noted earlier, all Indian languages are more flexible with regard to word order and allow for basically any constituent to be fronted. Furthermore,

they have the basic word order SOV. This may be transferred onto Indian English in the sense that the placing of constituents (other than the subject) at the beginning of the clause is less marked, especially the placing of objects before the verb. In other words, speakers of Indian English structure their sentences according to the principles they are familiar with from their L1 and put elements in clause-initial position, which would normally not occur there, more frequently than speakers of other varieties because the clause-initial position is less marked for them. Fronted objects are particularly common, which suggests that the word order OSV seems more natural to Indian English speakers or may even be part of their repertoire of basic sentence patterns. It is rather unlikely, however, that the OSV sentence structure will eventually replace the SVO sentence structure in Indian English. That is, it is unlikely that we are here dealing with a case of word order change, a process which is in fact not uncommon in language contact situations (cf. Heine/Kuteva 2003; Heine 2008). Rather, the findings of the present analysis indicate that Indian English speakers are more relaxed than the speakers of the other varieties of English analyzed as far as word order is concerned.

4.4 Existential *there*-constructions

This section examines existential *there*-constructions, sentences that begin with the word *there*, followed by a form of *be*, which in turn is followed by a noun phrase (the notional subject), as in *There's no food in the fridge*. Typically, these constructions are used to express the existence or occurrence of some entity at some place. The general overview at the very beginning of this chapter has already indicated that existential clauses are used much more frequently than the other information-packaging constructions analyzed in this study. Furthermore, their distribution across the varieties of English analyzed is more balanced than that of the other constructions. The following subsections will examine existential clauses in some more detail. The discussion of the overall distribution will be followed by a comparison of bare and extended existentials and the investigation of variable concord in existentials with plural notional subjects.

4.4.1 Overall distribution

Figure 4.14 plots the distribution of existential *there*-constructions in the nine varieties of English analyzed. As can be seen, the construction is most frequently used by speakers of Irish English and British English, followed by speakers of Indian English (369.9, 338.2 and 297.7 tokens per 100,000 words, respectively). The frequencies between these varieties are not significantly different ($p > 0.05$ in chi-squared tests), but the difference between Irish English and all the other varieties is statistically highly

significant ($p < 0.001$).

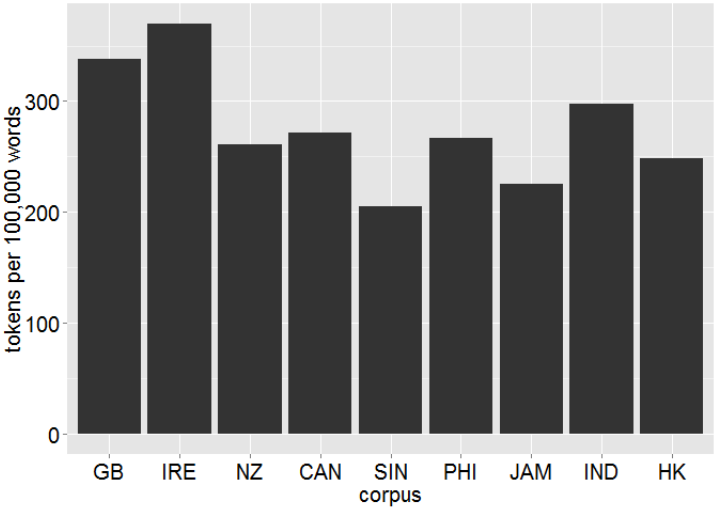

Figure 4.14: Existential *there*-constructions in the S1A-files of nine ICE corpora (frequencies per 100,000 words).

Singapore English and Jamaican English show the smallest frequencies of existential clauses among the varieties analyzed (205.1 and 225.5 tokens per 100,000 words, respectively).[38] The data suggest that this may be due to the fact that speakers of these two varieties use alternative constructions more frequently to express existence than do the speakers of the other varieties. In both varieties these alternative existential constructions involve possessive verbs, *got* in Singapore English and *have* in Jamaican English. The use of possessive verbs to express existence is not unique to these two varieties of English, but it is in fact the case in many languages of the world which, according to Her (1991: 383), is "largely because the relation of possession between two entities is in fact a kind of relation of existence as well". Freeze (2001) also discusses the relation between possession and existence. He places existential *there*-constructions within the so-called 'locative paradigm', assuming that existentials are "universally locative" (2001: 941).

Turning to Singapore English first, the verb *got* has assumed a number of new functions, including that of expressing existence (Lee et al. 2009; Bao 2014). The following examples illustrate this function of *got*, which is widely used in colloquial speech. In (4.66), the speakers talk about an open air festival with music and theatre performances. The three examples of existential *got* are marked in bold print.

(4.66) A: There's a lot of fun there's a lot of fun
B: Uhm
A: Very good lah

38 In Singapore English, the frequency is significantly smaller than in all other varieties (at $p < 0.05$ for Hong Kong English; at $p < 0.01$ for New Zealand, Canadian, Philippine English; at $p < 0.001$ for British, Irish and Indian English) but Jamaican English ($p > 0.05$).

B: So the people who were trained are youngsters

A: **Got some *ang mos***

B: **Got a lot of *kwai lo* uh**[39]

A: No there are more there are more locals than there are *ang mos*

B: Huh

A: Then **got some very yuppie yuppie types**

(ICE-SIN:S1A-025)

The three sentences with existential *got* are all examples of Chinese-style existential quantification, a special case of existential *got*. They can be paraphrased as *There are some ang mos*, *There are a lot of kwai lo* and *There are some yuppie types*, respectively. The construction is more likely to occur in colloquial speech, an impression underlined by the fact that the speakers use the terms *ang mo* and *kwai lo*, two derogatory terms rather used in colloquial speech.

Furthermore, existential *got* can be used to assert the existence of an entity at some location, as the examples in (4.67) illustrate. In this excerpt, two women praise a friend's fruit cakes. Audrey seems to be a little child also present during but not relevant for the recordings.

(4.67) B: And then she will bake all this type of fruit cakes you know. Ah there is one [unclear word] is very nice.

A: Huh.

B: [unclear word] cake **inside got fruits**.

A: A no Audrey better come down.

B: Different kind of **Inside got a lot of fruits**.

(ICE-SIN:S1A-006)

The sentences with existential *got* can be paraphrased as *There is (a lot of) fruit in the cake*, that is, *got* is here used to assert the existence of fruit in the cake (whose name the transcriber obviously had problems to understand). Prior to this piece of dialogue the women talk about other types of cake, producing further sentences with existential *got*, for example, *A apple pie inside got wine or not?*

Researchers agree that the novel uses of *got* in Singapore English are substrate influenced, derived from the Chinese construction marked by *u* in Hokkien, *yau* in Cantonese and *you* in Mandarin, which all mean 'have' and are used to express possession and existence (e.g. Her 1991; Lee et al. 2009; Bao 2014). Assuming that the functional properties of the construction are the same in the three Chinese dialects, Bao (2014) decides to cite Mandarin data in his paper. He suggests to approach the issue from a set-theoretic perspective and argues that "the set of morphosyntactic frames of *got* in Singapore English is derivable through the merger of the frames of English *get* and those of Chinese *you*" (2014: 152). The following Chinese examples

39 The terms *ang mo* and *kwai lo* are used by Chinese to refer to Caucasians.

give convincing evidence that the existential *got*-construction is derived from the Chinese *you*-construction.[40] The sentence in (4.68) is an example of *got* expressing existence, (4.69) is an example of existential quantification.

(4.68) *wu* *li* *you* *ren*
house in *got* person
'There are people in the house.'

(4.69) *you* *ren* *zhao* *ni*
got person look-for you
'There is a person who is looking for you.'

(Bao 2014: 151)

As can be seen in these two examples, the source of Singapore English existential *got* is very likely Chinese *you*, from which *got* inherits its uses.

Jantos (2009) has systematically analyzed possessive sentences with *have* and the subject pronouns *they* and *you* in ICE-Great Britain, ICE-Jamaica, ICE-India and ICE-Singapore and comes to the conclusion that possessive sentences with the subject pronoun *they* are most frequent in Singapore English (112 tokens as opposed to 56 tokens in British English, 66 in Indian English and 74 in Jamaican English; cf. Jantos 2009: 176). Furthermore, there is a prevalence of the strings *they got a* + noun phrase and *you got a* + noun phrase in Singapore English (Jantos 2009: 180). Regarding the subject-less type (i.e. *Got some ang mos*), Jantos finds 46 instances in the Singapore English data and only one or two in the other varieties.

Given all these findings, it is plausible to assume that existential *there*-constructions are less frequent in ICE-Singapore than in the other corpora analyzed because Singapore English speakers use constructions with existential *got* or possessive *have* more frequently to express existence than do speakers of the other varieties.

Turning to Jamaican English, substrate influence also seems to affect the expression of existence in this variety of English. Jantos (2009), in her analysis of possessive structures, notes that Jamaican English "has a strong preference for expressing existence by means of the possessive collocation *you have a* + noun phrase" (2009: 179). This is very likely due to influence from Jamaican Creole, where existence is normally expressed by means of possessive structures with the invariant verb *(h)av* plus an indefinite pronoun subject, for example, *yu* 'you' and *dem* 'they' (Patrick 2007: 140). Further evidence in support of this claim comes from the fact that Jantos finds the construction *them have* + indefinite noun phrase only in her Jamaican English data, which can very likely be attributed to influence from Jamaican Creole *dem (h)av* 'they have'. Consider the following example from Jamaican Creole for illustration of the existential construction.

(4.70) *dee* *hav* *a* *gruup* *a* *man* *niem* *Stepaz* *an* *yu* *hav* *dis* *poliis*
3P exist IND group of man name S and 2S exist DEM police
gai *we* *badigyaad* *Siyaga*
guy REL bodyguard S

40 Contrary to Bao (2005, 2014), Lee et al. (2009) argue that Singapore English *got* is derived from Hokkien *u* rather than Mandarin *you* or Cantonese *yau*.

'There was a group of men called the Steppers, and there was a policeman who bodyguarded Seaga.'

(Patrick 2007: 140)

This sentence shows that English *there was* may be realized in different ways in Jamaican Creole: in the first clause we have the verb *hav* plus pronoun *dee* 'they' and in the second clause it occurs together with the pronoun *yu* 'you' to express existence.

In sum, in Singapore English and Jamaican English substrate influence leads to the development and use of existential structures other than the *there*-existential construction, which, in turn, may account for the low frequency of the latter construction in these two varieties.

4.4.2 Bare vs. extended existentials

This section deals with the distinction between bare and extended existential clauses. Recall that bare existentials contain only *there*, the verb *be* and the notional subject (e.g. *There's a lot of fun*; ICE-SIN:S1A-025). Extended existentials, on the other hand, contain some additional material after the notional subject that is of relevance for the construction. These extensions include adverbials of time and place, as in (4.71a), predicatives, as in (4.71b), infinitivals, as in (4.71c), participials, as in (4.71d), and relative clauses, as in (4.71e).[41]

(4.71) a. There's a hole **in the tile**. (ICE-IND:S1A-040)

b. There's nothing **wrong** in that. (ICE-IND:S1A-049)

c. There is a good place **to sit**. (ICE-IND:S1A-032)

d. There's snow **coming**. (ICE-IRE:S1A-097)

e. There's a man **who watches his films**. (ICE-IRE:S1A-071)

Before turning to the distribution of these different types of extension, consider first of all the distribution of bare and extended existentials across varieties of English, presented in Table 4.15. As can be seen, bare existentials are the preferred variant of Indian English speakers (58.6% of all existential clauses), Philippine English and Singapore English speakers (53.5% each), which is not particularly surprising since learners of English usually tend to favour less complex constructions. Somewhat unexpectedly, however, Hong Kong English speakers and Jamaican English speakers use extended existentials more frequently than the bare variant (51.6% and 60.9%, respectively). The proportion of extended existentials in Jamaican English is in fact

41 For more details on the nature of existential *there*-constructions and the distinction between bare and extended variants see section 3.2.4. There, it is also pointed out that existentials can have two different types of participial clauses which must be distinguished, namely clauses that are an extension of the *there*-construction and clauses that are a modifier of the notional subject: *There's a festival starting this week in Derry* versus *There's a word beginning with D that would describe it.* Consequently, the category of participial extensions contains only the former type of construction.

the largest among all nine varieties analyzed. Recall that the investigation of LD and fronting constructions yielded similar results, with the L1 English speakers and Jamaican English speakers using more complex structures than the other L2 English speakers.

Table 4.15: Existential *there*-constructions in ICE (token frequencies and normalized frequencies per 100,000 words) plus the distribution of bare and extended variants (token frequencies and percentages out of all existentials).

corpus	existential		bare		extended	
	N	norm.	N	%	N	%
GB	682	338.2	314	46.0	368	54.0
IRE	746	369.9	313	42.0	433	58.0
NZ	600	260.9	248	41.3	352	58.7
CAN	573	272.0	259	45.2	314	54.8
SIN	417	205.1	223	53.5	194	46.5
PHI	578	266.8	309	53.5	269	46.5
JAM	481	225.5	188	39.1	293	60.9
IND	643	297.7	377	58.6	266	41.4
HK	591	248.3	286	48.4	305	51.6

The high frequency of extended existentials in Jamaican English is mainly due to relative clause extensions, which account for more than 57% of all complex existentials, as can be seen in Figure 4.15. The figure plots the proportions of the different types of extension in percentages out of all extended existentials.[42]

Existentials with relative clause extensions are also quite common in the other corpora, accounting for more than 30% of all extended existentials. In ICE-Ireland, however, they make up only around 20%. Interestingly, a similarly deviant behaviour of the Irish English component is attested for the subordinator types in IT-clefts, discussed in section 4.5. In IT-clefts, Irish English speakers use *wh*-forms and *that* less frequently as subordinators than the speakers of the other varieties of English analyzed. Zero subordinators are most common and *ing*-complements are also quite frequently chosen (most frequent after New Zealand English). In the case of existential clauses, Irish English even shows the highest proportion of participial extensions among the varieties of English analyzed (20.3%). The class of participial extensions includes here both present and past participles, but nearly 74% of the participles are of the former type in Irish English. An increasing use of *ing*-complements and a broadening of its functional range is reported in the literature to be the case in English generally (e.g. De Smet 2013; Duffley 2000; Fanego 1996, 2007; Mair 2002a, 2013; Rudanko 1998, 2000; among many others), but the results of the present analysis suggest that Irish English is more advanced in this development.

It is worth having a closer look at relative clause extensions because different preferences in the choice of relative pronoun/particle can be attested among the speakers

42 These proportions and the corresponding token frequencies are provided in Appendix 6.9.2.

of the nine varieties of English analyzed. The distribution of relative clause extensions with zero marking is of particular interest. They occur along with *wh*-forms, which are normally regarded as the more formal variant mainly used with human antecedents in both restrictive and non-restrictive clauses, and invariable *that*, acceptable only in restrictive clauses.

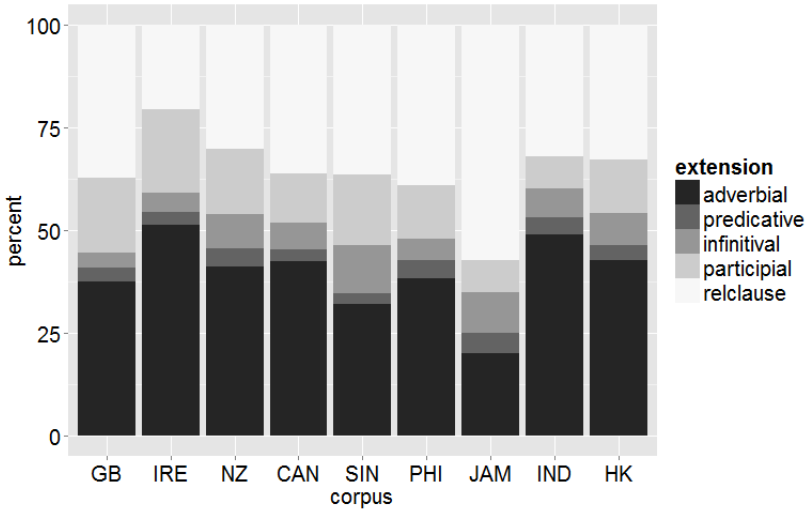

Figure 4.15: The type of extension in complex existential *there*-constructions (percentages out of all extended existentials).

Consider the sentences in (4.72) for illustration of the different types. While sentences (a) and (b) have the explicit relative markers *who* and *that*, respectively, the sentences in (c) to (f) are all instances with zero relatives. In (c), we have the subject *you* in the relative clause, while the other three examples are instances of subject relative clauses. What distinguishes the last three sentences is that the verb can either be tensed (present or past), as in (d) and (e), or untensed, as in (f).

(4.72) a. There're so many people **who need physiotherapy** (ICE-GB:S1A-003)

b. There is energy **that you can get from the sun** (ICE-JAM:S1A-089)

c. There's not much **you can do about it** (ICE-GB:S1A-006)

d. And there's this girl **comes in from another school and wins a dance competition** (ICE-IRE:S1A-012)

e. There's cardboard tubes **just fell down** (ICE-IRE:S1A-092)

f. But <,> is there is there somebody **pick you up** (ICE-HK:S1A-074)

Subject-zero relatives are today usually used only in existential *there*-constructions and *it*-clefts (Fischer 1992: 307), but they are reported to be a widespread phenomenon in these environments in a number of varieties of English. Lodge (1979) describes their use in a British dialect spoken near Manchester. Preusler (1938, 1942), Harris (1993) and Filppula (1999) discuss the omission of subject relative markers in Irish

English. While Preusler (1938) explains the high incidence of zero relatives in terms of Celtic influence, Filppula is more hesitant to ascribe full responsibility to Celtic as the source of this specific feature because it is so widespread in other varieties as well, he argues.

Li (2000), Hung (2012) and Newbrook (1988, 1998) describe zero relatives as a common feature of Hong Kong English speakers, even among highly educated ones. Newbrook (1988) describes it as the most common "error" in students' writings when using relative clauses. As a possible explanation for this observation he suggests that "the error is so common locally that most students would almost automatically produce it" (Newbrook 1988: 31) and adds that most students have probably never been taught that subject relatives can never be omitted.

Schachter and Celce-Murcia (1977) also report zero relatives as a frequent feature in the writings of Chinese learners of English. They argue that the high incidence of the construction is due to the fact that the Chinese students attempt to structure English sentences according to the principles of their topic-prominent L1. That is, they establish a topic first and then say something about it. Ortega (2009) takes up their ideas and suggests that existentials with subject-zero relatives are part of a developmental continuum from most L1-like to most L2-like information structure. In this developmental continuum, left dislocation constructions are at the most L1-like end, followed by existentials with zero-subject relatives. The most L2-like end of the continuum is taken by the canonical SVO sentence.

Christie (1996), examining Jamaican relative clauses, notes that subject-zero relatives are "the preferred alternative where the head noun belongs to a nominal complement represented by an indefinite NP, and the relativized NP represents the subject of the clause" (1996: 55; the sentences in (4.72d-f) fall into Christie's category).[43] She adds that this is best illustrated by an existential clause:

(4.73) *Dier wozz a liedu liv wid tuu children*

'There was a lady (who) lived with two children.'

(Christie 1996: 55)

Subject-zero relatives are also reported to occur frequently in Jamaican Creole (Patrick 2012).

In order to test these various claims on the frequency of subject-zero relatives in different varieties of English I counted all the instances of *wh*-forms, *that* and zero in subject and nonsubject position in existentials with relative clause extensions. The results can be seen in Figure 4.16. Since I am particularly interested in subject relative clauses, objects, adverbials, complements and possessives are grouped together as nonsubjects. The x-axis in Figure 4.16 presents for each ICE corpus the percentages of *that*, *wh*-forms and zero out of all subject relatives; the y-axis presents the percentages of the three variants out of all nonsubject relatives. That is, if a corpus label occurs more towards the lower left-hand corner, it is infrequently realized as the relevant variant in both subject and nonsubject position. If it occurs towards the

43 In Christie's (1996) study the term 'Jamaican' comprises "a range of varieties, all of which are seen as distinguishable from internationally accepted Standard English, although using a lexicon mainly derived from English. Some diverge more radically from Standard English than others" (1996: 48).

upper left-hand corner, the relevant option often occurs in nonsubject position but not in subject position. Labels that are on or near the diagonal (angle bisector) occur in subject and nonsubject position to roughly the same extent.

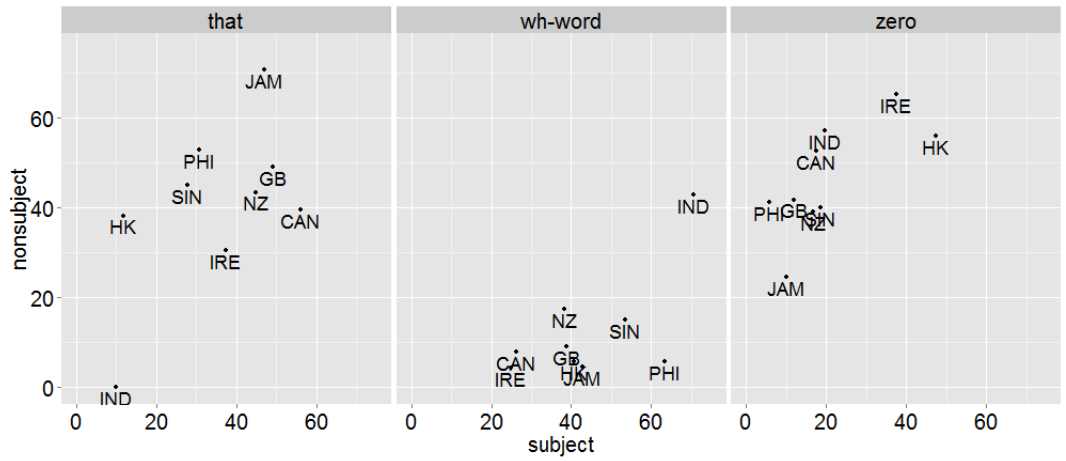

Figure 4.16: Relative clause extensions with *wh*-words, *that* and zero as relative marker in subject and nonsubject position (percentages out of all subject and nonsubject relative clauses, respectively).

As for zero relatives, all labels occur towards the upper left-hand corner, which means that this variant makes up a greater proportion among the nonsubject relatives than among the subject relatives in all varieties analyzed. The findings for Irish English and Hong Kong English are remarkable in two ways. First, they show the highest proportions of zero marking in both subject and nonsubject position among the varieties of English analyzed (the two labels are the ones furthest to the right and furthest to the top). And second, zero relatives are the most frequent option in subject position within both corpora, in Irish English together with *that* (47.5% of subject relatives in the Hong Kong data and 37.5% in the Irish data). These findings underscore the claims of previous studies arguing that zero relatives are common in these two varieties of English.

What is furthermore interesting to note is that *that* is the preferred variant in subject relative clauses among the speakers of the four L1 English varieties (together with zero in Irish English), whereas in the L2 varieties *that* occurs more frequently in nonsubject position. The latter tend to use *wh*-words in subject position (or zero in the case of Hong Kong English). Jamaican English is an exception in this respect, with *that* being the first choice in all positions. The findings for the L1 varieties are in line with previous studies on the variation of relative markers, reporting an increase of *that* (and zero) at the expense of *wh*-forms in spoken English (e.g. Tottie 1997; Biber et al. 1999; Tagliamonte 2002). The discrepancy in the use of *that* between the speakers of the L1 varieties (and Jamaican English) and the L2 varieties may be accounted for by the fact that the former are more relaxed as far as the animacy rule is concerned. That is, the former may be more relaxed in using *that* after a human antecedent while the latter stick more rigidly to the rule that a *wh*-form should be

used in such cases. This seems to be particularly true for speakers of Indian English, as the outliers IND in the facets 'wh-word' and 'that' suggest.

Returning to the claims about Hong Kong English and L1 Chinese learners of English mentioned above (Schachter/Celce-Murcia 1977; Newbrook 1988, 1998; Li 2000; Ortega 2009; Hung 2012), the present study shows that Hong Kong English speakers indeed use zero relatives most frequently in subject position among the varieties of English analyzed. The sentences that occur in the direct conversation files in ICE-Hong Kong furthermore suggest that this is a learner feature because the sentences contain further features which clearly signal a lack of proficiency. Consider the sentences in (4.74) for illustration.

(4.74) a. And and and there's a in few months ago there's a uhm Legislative councillor come to my school uh to give a speak *lauh waih hing* (ICE-HK:S1A-062)

b. There are many many of friend go by train went by train (ICE-HK:S1A-062)

c. In my school <,> there's so many student like uh bad behave (ICE-HK:S1A-031)

d. But I I'm worry about at the present time whether <,> a man can <,> uh <,> whether there is a man <,> will marry with me (ICE-HK:S1A-054)

These sentences clearly reveal that the speakers are learners of English. Examples (a) and (b) are by the same speaker. He or she seems to have problems with tense marking. Furthermore, the word *speak* is used instead of *speech*. Also note the code-switching in sentence (a). In (c), plural and adverb markings are missing and the adverb *bad* should actually occur after the verb. And finally, the whole structure of sentence (d) is confusing. Furthermore, the preposition *with* is inserted after *marry*.

The constituents in these sentences appear to be individual chunks, only loosely connected. That is, the relative clause extensions do not seem to be really part of the existential *there*-constructions, but they rather seem to be constituents on their own. These sentences can quite plausibly be regarded as attempts on the part of the speakers to structure their sentences according to the principles of their topic-prominent L1, that is, establishing the topic first and then saying something about it (cf. Schachter and Celce-Murcia 1977; Ortega 2009). This may well be the primary motivation for Hong Kong English speakers to use such constructions. A somewhat lower proficiency level in comparison to Singapore English speakers may make them more susceptible to transfer of the information structure of the L1; or as Ortega (2009) puts it: "[w]ith time and increasing proficiency, the tendency to transfer the information structure of the L1 in order to frame ideas in the L2 may gradually diminish" (2009: 46).

Given these findings, the model of the developmental continuum suggested by Ortega (2009) seems to be a plausible way of accounting for the interaction between a speaker's proficiency level, the degree of influence from the L1 on the L2 and the preferred information structure. Recall that left dislocation constructions constitute the most L1-like end of the continuum because the information structure of these sentences is very much like the basic sentence structure of topic-prominent languages

(e.g. *Many people, they find her okay*). Existentials with subject-zero relatives constitute the second most L1-like stage (e.g. *There are many people find her okay*), followed by existentials with overt relative markers (e.g. *There are many people who find her okay*). The most L2-like end of the continuum is taken by canonical SVO sentences (e.g. *Many people find her okay*). Since left dislocation constructions are not particularly more frequent in Hong Kong English than in the other L2 English varieties analyzed, it may well be the case that the Hong Kong English speakers included in the ICE corpus are already too advanced as far as their proficiency of English is concerned to use left dislocation more frequently.

Singapore English also ties in with the argument. We find only a small number of existentials with zero-subject relatives and the smallest number of LDs among the L2 English varieties analyzed. This may be due to the fact that Singapore English speakers are more proficient in English and thus transfer the information structure of their L1 less frequently.

The question that now arises is how does Irish English fit in with this argument and the developmental continuum because existentials with subject-zero relatives are also quite common in this variety of English. As the sentences in (4.75) illustrate, in Irish English existentials with subject-zero relatives are much more compact constructions than is the case with those items attested for Hong Kong English. That is, the *there*-clause and the relative clause form one construction rather than individual chunks.

(4.75) a. There's not so many people knit now (ICE-IRE:S1A-045)

b. There's only two dentists in the Newry and Mourne area are girls (ICE-IRE:S1A-022)

c. Is there any residents like that are always there (ICE-IRE:S1A-014)

d. There's cardboard tubes just fell down (ICE-IRE:S1A-092)

A combination of several reasons may be responsible for the higher use of zero relatives in both subject and nonsubject position in Irish English as compared to the other L1 English varieties in particular. It is quite likely that Celtic influence played an important role in the establishment of zero relatives in earlier English and that the constant presence of Irish in the contact ecology of Irish English made the construction more entrenched in the speech of Irish English speakers than in that of, say, British English speakers. Consider the following examples of zero relatives given by Preusler (1938) to show the Celtic source of the feature. Note that the Celtic languages are VSO languages which means that the unmarked counterpart to the sentence *John is strong* would be *y mae John yn gryf*, 'is John strong'. In order to put emphasis on John or to introduce him as the one who is strong, the construction in (4.76) is used.

(4.76) *John sydd yn gryf*
John is strong
'It is John (who) is strong.'

(4.77) *pwy sydd yma*
who is there
'Who is it (that) is there?'

(Preusler 1938: 184)

Unfortunately, Preusler gives no examples of existential clauses with zero relatives, but the mechanisms of the Irish IT-cleft seem to be readily transferable to existential *there*-constructions.

Furthermore, the high incidence of zero relatives can at least in part be explained in terms of diffusion from earlier stages of English because zero relatives were much more common in Old English and Middle English in both subject and nonsubject position (cf. e.g. Fischer 1992). The retention of early dialectal features of British English in the speech of Irish English speakers has already been attested for other characteristic features of this variety of English and thus seems to be a plausible explanation as well (e.g. Filppula 1999).

The preference of zero relativizers, which Christie (1996) notes for Jamaican relative clauses, cannot be attested for the Jamaican English speakers included in ICE because the proportion of zero relatives in subject relative clauses is much smaller than that of *wh*-words and in particular *that*, as can be seen in Figure 4.16.

The previous discussion of extended existential clauses has shown that speakers of English prefer different types of extensions. Variation can also be attested in terms of singular or plural agreement with plural notional subjects, as the following discussion will show.

4.4.3 Singular concord

In English, existential *there*-constructions with plural notional subjects vary between plural concord (4.78a) and singular concord (4.78b):

(4.78) a. There are biscuits there (ICE-IRE:S1A-069)

b. There's little benches outside (ICE-IRE:S1A-089)

This variability has been investigated in many studies which provide interesting results on British English (Martinez Insua/Palacio Martinez 2003; Crawford 2005; Rupp 2005), American English (Schilling-Estes/Wolfram 1994), Canadian English (Meechan/Foley 1994; Walker 2007), Australian English (Eisikovits 1991) and New Zealand English (Britain/Sudbury 2002; Hay/Schreier 2004). More recently, variable concord has been analyzed in a number of ICE corpora (Jantos 2009; Collins 2012). The findings of these studies are hard to compare as they are based on different types of data and different definitions of the variables that are examined. Yet, despite this inconsistency they also reveal some common tendencies. The present study will build on and add to the valuable insights gained by the previous studies by examining variable concord across a range of English varieties and testing some of the factors identified in the literature as constraining singular concord.

Collins' (2012) study is very similar in its nature to the present one. It also investigates variable concord in the 'direct conversation' files of ICE-Great Britain, ICE-Singapore, ICE-Philippines, ICE-India and ICE-Hong Kong. Nonetheless, I think it is worth looking at variable concord again because the present study provides data of further corpora not included in Collins' study (ICE-Ireland, ICE-New Zealand, ICE-Canada and ICE-Jamaica). Furthermore, in contrast to Collin's approach, I think it

is necessary to also include the relevant proportions for plural concord. That is, the examination of the factors identified in the literature as influencing singular concord should also consider the cases where speakers use plural concord in existentials with plural notional subjects. This is important for the following reason: say, a variety of English shows a high proportion of contracted forms of *be* in existentials with singular concord; in order to be able to conclude that there is a correlation of singular concord with contraction it is necessary to also know the number of incidences of plural concord with contracted forms of *be* because it may well be the case that in this specific variety of English contracted forms are generally more prevalent than in other varieties. Hence, the present study examines the linguistic factors tense, contractedness, complexity, polarity and type of determiner (before the notional subject) in both cases, singular concord and plural concord.

Before turning to the distribution of singular concord across the varieties of English analyzed, note that in some L2 English varieties the relation between determiner choice and number of the notional subject may at times be confused, as in the sentence in (4.79a) taken from the Indian English sample, where we find the indefinite article *a* before the plural noun *differences*. Furthermore, sometimes the number of the notional subject is not consistently marked, as in (4.79b) and (4.79c), where we find *university* and *examination*, while the prenominal elements suggest that plural nouns be used.

(4.79) a. [...] and if so many parties come together <,> uh there is a ideological differences <,> (ICE-IND:S1A-005)

b. But in Hong Kong there are only three three or six six university now so (ICE-HK:S1A-030)

c. I think uhm <,> there's <,> many examination (ICE-HK:S1A-060)

The omission of plural marking is particularly frequent in the Hong Kong English data of the present study. This finding is in line with previous research, describing the omission of plural marking as a common feature of this variety of English (cf. e.g. Budge 1989; Setter et al. 2010). Notional subjects, such as the ones in sentences (b) and (c), are treated as plural in the present analysis. In cases such as the one in (a), the context has carefully been considered to determine whether the notional subject is really plural.

Let us now turn to the distribution of existentials with singular concord across the varieties of English analyzed. Figure 4.17 plots the proportions of singular concord out of all existentials with plural notional subjects.[44] [45]

In line with previous cross-varietal studies, the L1 English varieties are more accepting of singular concord than the L2 varieties (e.g. Jantos 2009; Collins 2012). This is very likely due to the fact that in the latter cases English is usually associated with formality and spoken in more formal settings, while the local languages are used

44 Note that there is some variation with respect to the use of plural noun phrases as opposed to singular ones across the varieties of English analyzed. While the proportion of plural noun phrases amounts to only 23.6% in the Canadian English data, it makes up 45.4% in the Hong Kong English data. In the other data the proportions range from 27.7% to 35.2%.

45 The exact percentages and token frequencies are given in Appendix 6.9.2.

in informal situations, those situations where singular concord usually occurs. Put differently, L2 English speakers do what they have learned at school, namely use a plural verb form together with a plural noun. What is surprising, however, is the vast discrepancy between the frequencies of singular concord in the three L1 English varieties Irish English, New Zealand English and Canadian English on the one hand and the other varieties on the other, with percentages of singular concord amounting to 73.4% in the Irish English sample and only 9.5% in the Indian English one.

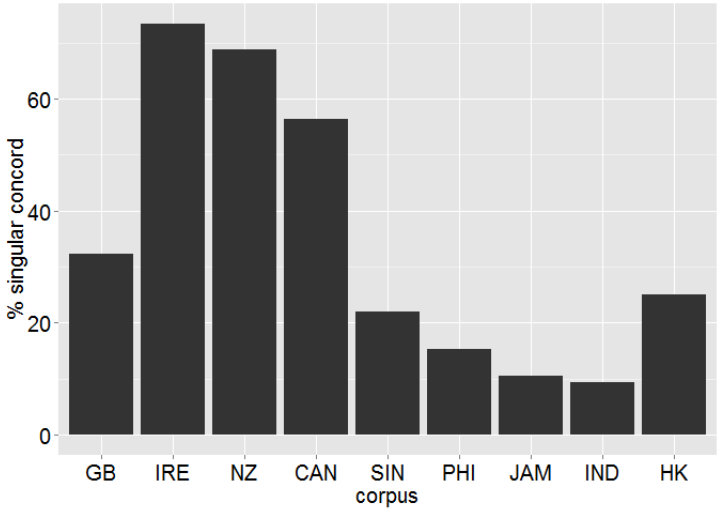

Figure 4.17: The proportions of singular concord out of all existential *there*-constructions with plural notional subjects.

The small proportion of singular concord in ICE-Great Britain as opposed to the other L1 English corpora may be due to formality reasons since the direct conversation files in the British component have a higher level of formality than the conversations included in the other L1 corpora (cf. the discussion of the data in the introductory chapter of this study). Evidence in support of this assumption is also provided by Martinez Insua and Palacios Martinez (2003), who find a higher proportion of singular concord in the British National Corpus (BNC) than in ICE-Great Britain in both the written texts (3.22% in BNC vs. 1.39% in ICE) and the spoken texts (13.26% in BNC vs. 6.41% in ICE). Note that the BNC is commonly described as containing more informal data than ICE-Great Britain.

Among the L2 English varieties, Hong Kong English speakers are most accepting of singular concord, which is somewhat surprising although in line with Collins' (2012) findings. Collins argues that "perhaps the persistence of the type of informal local elements that are the butt of the strong complaint tradition in Hong Kong provide a nurturing environment for SA [i.e. singular agreement; cw]" (2012: 63). To this I would add that plural marking on nouns is not consistently present in Hong Kong English, as noted earlier. This in turn may lead to a confusion in verb-noun agreement, with Hong Kong English speakers simply not paying that much attention to it or not giving much importance to it.

In order to better understand the distribution of singular concord across the varieties analyzed the following paragraphs will examine a number of factors that are reported in the literature to covary with singular concord. The factors that will be investigated include tense, contractedness, extension, polarity and the type of determiner preceding the notional subject. The factor tense distinguishes between present tense (4.80a) and past tense (4.80b)[46]; the factor contractedness compares contracted forms (4.80c) with full forms (4.80a-b).

(4.80) a. There is contingency plans (ICE-IRE:S1A-024)

b. There was four people in it (ICE-IRE:S1A-009)

c. There's little benches outside (ICE-IRE:S1A-089)

Most studies report singular concord to covary with present tense rather than past tense (e.g. Britain and Sudbury 2002; Eisikovits 1991). However, taking the interaction between tense and contractedness into account (almost all contracted forms occur in present tense contexts), Hay and Schreier (2004) and Walker (2007) note that past tense increases the likelihood of singular concord for full forms. That is, if we consider only full forms singular concord covaries with past tense rather than present tense. As for the factor contractedness, contracted forms are reported to promote singular concord more than full forms do (Hannay 1985).

The factor extension investigates the effect of bare (4.81a) as opposed to extended plural existentials (4.81b) on concord. Many studies show that extended existentials are more strongly associated with singular concord than bare instances (e.g. Hannay 1985; Martinez Insua/Palacios Martines 2003; Walker 2007).

(4.81) a. There was videos (ICE-IRE:S1A-046)

b. There was twenty of them going for dinner in where's it now in Temple Bar (ICE-IRE:S1A-049)

The factor polarity distinguishes between existentials with negative and positive polarity. Negative existentials include instances with the negative particles *no* and *not*, as in (4.82a) and (b), the latter also in contracted form, as in (c). Many studies find negatives to favour singular concord (Meechan/Foley 1994; Martinez Insua/Palacios Martinez 2003; Rupp 2005), though others find the opposite (Britain/Sudbury 2002).

(4.82) a. There was no cars (ICE-IRE:S1A-079)

b. There's not enough partners (ICE-NZ:S1A-029)

c. There wasn't too many (ICE-NZ:S1A-017)

46 Note that the data also include examples where *be* is preceded by a modal auxiliary, as in the following examples.

a. There **must** be buses going along Mansfield Road (ICE-GB:S1A-023)

b. There **will** be sixteen or seventeen people (ICE-HK:S1A-053)

The notion of concord is, of course, of no relevance in these sentences and consequently they are excluded from further consideration in the present study.

All *there*-existentials with plural notional subjects have been annotated according to these four factors - tense, contractedness, extension and polarity. Additionally, the type of determiner preceding the notional subject has been examined. This factor will be discussed separately below because of several reasons. First, the factor did not emerge as significant in any of the samples analyzed and, second, since there are seven different types of determiner to be distinguished some subcategories contain only very small numbers of tokens and the results are therefore not very reliable.

The proportions of the factors tense, contractedness, extension and polarity in existentials with plural and singular concord are plotted in Figure 4.18 and Figure 4.19 below. The x-axes present the proportions in existentials with singular concord, the y-axes those with plural concord.[47]

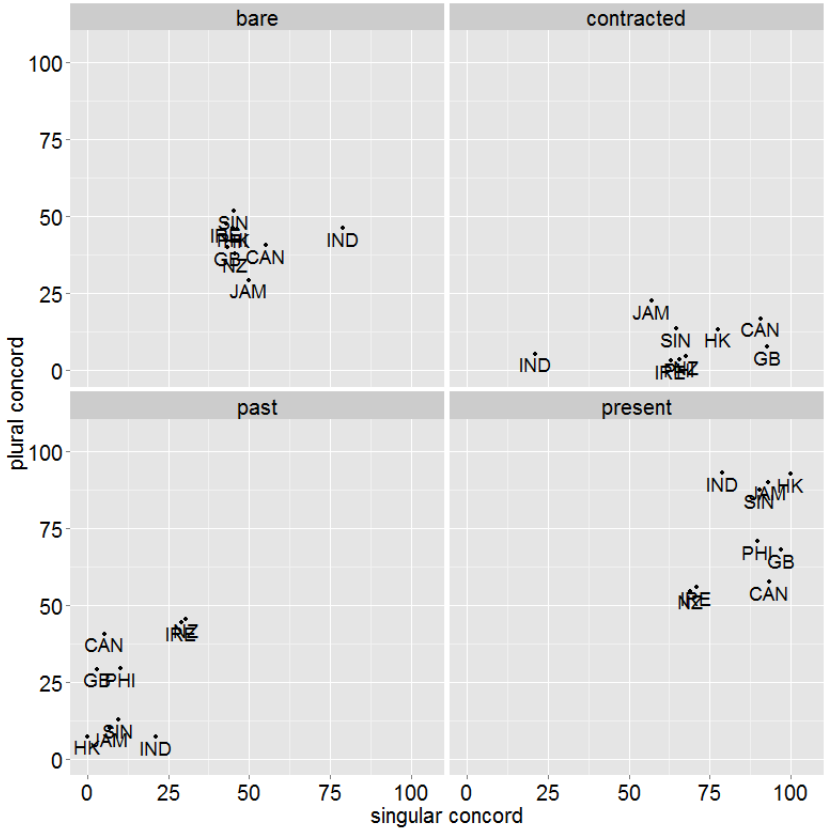

Figure 4.18: Effects of the factors extension, contractedness and tense on singular concord and plural concord (x-axis: percentages out of all existentials with singular concord; y-axis: percentages out of all existentials with plural concord).

Interestingly, the labels for each factor are roughly grouped together, that is, the labels for bare existentials are all positioned in the middle of the graph, the labels representing contracted forms are located towards the lower right-hand corner, the labels representing past tense forms are located towards the lower left-hand corner

47 The raw frequencies and exact percentages are given in Appendix 6.9.2.

and the labels in the facet 'present' are positioned towards the upper right-hand corner. This means that the proportions of the factors are roughly the same across the samples. The plots can be interpreted in the following way. First, in all nine varieties of English analyzed the factor most strongly associated with singular concord seems to be contraction, and especially so in British English and Canadian English, where over 90% of existentials with singular concord covary with contracted forms as opposed to less than 20% of existentials with plural concord doing so. Indian English is the outlier in this respect, but note that contracted forms are generally rarely used in this variety of English (the percentages are very low for existentials with both plural concord and singular concord).

As far as the factor tense is concerned, it can be noted that present tense forms are generally used more frequently than past tense forms in both types, existentials with singular concord and those with plural concord (in the facet 'present', all labels are located in the upper right-hand quadrant, whereas in the facet 'past' all labels are positioned towards the lower left-hand corner).

Regarding the factor extension, it can be noted that in Indian English singular concord is strongly associated with bare existentials, accounting for 80% of all existentials with plural notional subjects. A similar tendency can be attested for Canadian English (56% of bare existentials) and Jamaican English (50%), but here the covariation is not as pronounced. In all other varieties the proportions of bare and extended existentials with singular concord are much more balanced.

The facet 'negative', represented in Figure 4.19, shows that existentials rarely have negative polarity in both cases, existentials with plural concord and those with singular concord (all labels are positioned towards the lower left-hand corner). Negative polarity is slightly more common in both cases in the New Zealand English data, while it seems to favour singular concord in the Jamaican English data. In the latter sample, nearly 36% of existentials with singular concord have negative polarity, while less than 4% of existentials with plural concord are negative.

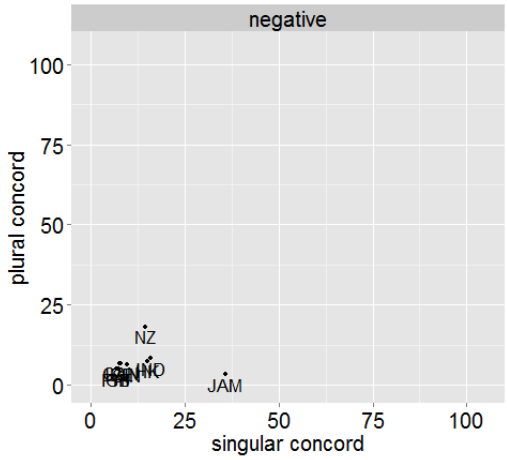

Figure 4.19: Effect of the factor polarity on singular concord and plural concord (x-axis: percentages of negative existentials out of all existentials with singular concord; y-axis: percentages out of all existentials with plural concord).

In order to test whether these intuitions are correct and to determine the presence and nature of any correlations, a binary logistic regression model was used.[48] Logistic regression estimates the probability of a given outcome (here: singular concord) given a number of predictors. This technique is widely used in sociolinguistics, where it is known as VARBRUL analysis. The function *lrm* offered by the *rms* package in R was used to estimate the probabilities of singular concord in the nine samples given the predictors tense, contractedness, extension and polarity. The results are presented in Table 4.16. Let us go through this rather large table step by step.

In the first column, the statistics listed below the name of the relevant variety samples assess the goodness of fit of the models. 'LR chi2' stands for model likelihood chi-square, the difference between the null deviance and the residual deviance. It is useful for ascertaining whether the predictors in the full model jointly earn their keep. As can be seen in Table 4.16, this measure is comparatively low in Jamaican English, Indian English and Singapore English, indicating that the explanatory value of the three models is not as good as that of the other models. But the p-values are still small (0.0007, 0.0019 and <0.0001, respectively), so we still have models with some explanatory value.

The C-index is a measure that addresses the predictive ability of the model. It is an index of concordance between the predicted probability and the observed response. That is, if C takes the value of 0.5 the predictions are random and they are perfect if C equals 1. A C-index above 0.8 is commonly regarded to indicate that the model has some real predictive capacity (Baayen 2008: 204). As can be seen in Table 4.16, the predictive abilities of the models are all quite good. The models with the best predictive abilities are the ones for the British English and New Zealand English samples (0.937 and 0.903, respectively). The accuracy and predictive capacity of the Indian, the Jamaican and the Singaporean models are not as good as those of the other models. This is very likely due to the small samples and, in particular, the very small number of existentials with singular concord. In the Indian sample there are only 19 items of existentials with singular concord, in the Jamaican sample 14 items and in the Singaporean sample 31 instances (cf. Appendix 6.9.2). This reminds us of the fact that all models are in any case imperfect simplifications. In the present case, it would certainly be good to back up the results by an analysis based on larger datasets.

Turning to the third column, it lists the estimates of the coefficients. The coefficient for TENSE=PRESENT expresses the contrast between present and past (the reference level mapped onto the intercept); that for CONTRACTION=FULL the contrast between full and contracted (reference level); that for EXTENSION=EXTENDED the contrast between extended and bare (reference level); and that for POLARITY=POSITIVE the contrast between negative (reference level) and positive polarity. In the Irish sample, for example, the negative coefficients for tense and contraction indicate that the probability of singular concord goes down with full forms and present tense contexts. That is, singular concord favours past tense contexts and contracted forms. This effect is statistically significant as the small p-values show (last two columns). The column named 's.e.' gives the estimated standard errors, which measure the accuracy with which the sample represents the population.

48 I am indebted to Florian Schirm who was of great help in doing the statistics.

Table 4.16: Effects in the logistic regression model for concord variation.

		coef	s.e.	p-value	
GB	intercept	3.2	1.48	0.033	*
LR chi2 = 158.13	tense=present	-0.38	0.94	0.688	
p < 0.0001	contraction=full	-5.24	0.70	<0.001	***
C = 0.937	extension=extended	-0.98	0.64	0.127	
	polarity=positive	-0.63	1.09	0.565	
IRE	intercept	5.15	1.10	<0.001	***
LR chi2 = 94.73	tense=present	-1.51	0.41	<0.001	***
p < 0.0001	contraction=full	-4.96	0.79	<0.001	***
C = 0.872	extension=extended	0.56	0.39	0.146	
	polarity=positive	0.08	0.80	0.922	
NZ	intercept	6.59	1.06	<0.001	***
LR chi2 = 121.7	tense=present	-3.37	0.78	<0.001	***
p < 0.0001	contraction=full	-6.53	0.96	<0.001	***
C = 0.903	extension=extended	0.58	0.44	0.187	
	polarity=positive	0.02	0.54	0.973	
CAN	intercept	3.35	1.49	0.024	*
LR chi2 = 84.33	tense=present	-0.98	0.89	0.276	
p < 0.0001	contraction=full	-4.45	0.82	<0.001	***
C = 0.893	extension=extended	-0.18	0.52	0.739	
	polarity=positive	-0.39	1.07	0.715	
SIN	intercept	0.93	1.11	0.405	
LR chi2 = 31.29	tense=present	-0.80	-1.08	0.282	
p < 0.0001	contraction=full	-2.60	0.51	<0.001	***
C = 0.790	extension=extended	0.21	0.47	0.653	
	polarity=positive	0.05	0.85	0.955	
PHI	intercept	2.52	1.19	0.034	*
LR chi2 = 61.81	tense=present	0.06	0.73	0.932	
p < 0.0001	contraction=full	-4.06	0.63	<0.001	***
C = 0.838	extension=extended	0.26	0.55	0.632	
	polarity=positive	-1.57	0.88	0.075	
JAM	intercept	2.23	1.62	0.167	
LR chi2 = 19.16	tense=present	-0.77	1.18	0.515	
p = 0.0007	contraction=full	-1.49	0.69	0.030	*
C = 0.786	extension=extended	-0.63	0.65	0.334	
	polarity=positive	-2.78	0.84	0.001	**
IND	intercept	1.68	1.18	0.155	
LR chi2 = 17.04	tense=present	-1.50	0.68	0.028	*
p = 0.0019	contraction=full	-1.80	0.71	0.011	*
C = 0.727	extension=extended	-1.43	0.60	0.017	*
	polarity=positive	-0.60	0.76	0.429	
HK	intercept	0.68	1.22	0.574	
LR chi2 = 92.04	tense=present	0.33	1.07	0.756	
p < 0.0001	contraction=full	-2.98	0.36	<0.001	***
C = 0.831	extension=extended	0.11	0.36	0.755	
	polarity=positive	-0.50	0.56	0.367	

significance levels: p = 0.05 ., p < 0.05 *, p < 0.01 **, p < 0.001 ***

Looking at the last two columns, the statistics corroborate the hypothesis that contractedness has an effect on concord. In line with previous studies, the results are highly significant at the level $p < 0.001$ for most varieties of English analyzed, with contracted forms favouring singular concord (the coefficients for CONTRACTION=FULL are negative in all nine samples). In Jamaican English and Indian English, the p-values are slightly higher, though still significant at the level $p < 0.05$. Follow-up studies should test whether these findings are reliable or whether the somewhat different results for these two varieties of English are simply an artefact of the small sample sizes.

The factor tense has an effect in Irish English, New Zealand English and Indian English (at the level $p < 0.001$ in the two former varieties and at $p < 0.05$ in the latter). In all three varieties of English singular concord covaries with past tense rather than present tense (again, the coefficients are negative), which is in line with Walker's (2007) findings.[49] Note that the results for Indian English need to be taken with a pinch of salt because the overall frequency of existentials with singular concord is very small, as noted earlier (19 tokens). Back-up for this finding is therefore clearly needed.

Since there is an interaction between contractedness and tense in the sense that almost all contracted forms occur in present tense contexts, it was tested whether contraction also had an effect if only existentials in present tense contexts were taken into account. The results turned out significant for all nine varieties of English analyzed. That is, contractedness has an effect on concord independently of tense. In a similar way it was tested whether tense had an effect if only full forms were considered. The results were significant for Irish English, New Zealand English and Indian English, exactly the same varieties that had also significant results with respect to the factor tense in the model with all predictors and all data, that is, tense has an effect on concord independently of contractedness. This led me to keep both factors - contractedness and tense - in the model.

The factor extension is selected as significant in Indian English, where we find a correlation of singular concord with bare existentials. This is in contrast to Walker (2007) and Martinez Insua and Palacios Martinez (2003) in whose studies singular concord is associated with extended existentials.

The factor polarity shows an effect in the Jamaican English sample, with negatives favouring singular concord ($p < 0.01$).[50] Hence, the results of the present study are in support of both previous studies that claim that polarity has an effect on concord and those that find no effect (cf. e.g. Meechan/Foley 1994, Martinez Insua/Palacios Martinez 2003 or Rupp 2005 versus Britain/Sudbury 2002). It seems that it depends on the variety of English the data are taken from whether polarity has an effect on

49 In the Indian English sample, tense interacts with polarity in the sense that there are no negative existentials that occur in past tense contexts. If the factor tense is considered only in existentials with positive polarity, it is still significant ($p < 0.05$), that is, tense has an effect on concord independently of polarity. Polarity has no effect if only existentials in present tense contexts are considered.

50 In the Jamaican English sample, there is an interaction between tense and polarity in the sense that there are no negative existentials that occur in past tense contexts. I therefore tested whether the factor polarity also had an effect if only present tense contexts were considered. The results are highly significant ($p < 0.001$). If the factor tense is considered only in existentials with positive polarity, it has no effect.

concord or not. Recall, however, that there is only a small number of existentials with singular concord in the Jamaican English sample and the results are therefore possibly not reliable.

In sum, the regression model confirmed many of the assumptions gleaned from the plots in Figures 4.18 and 4.19. While contractedness favours singular concord in all nine varieties of English analyzed, we find more variation with respect to the other factors. Tense has an effect on concord only in Irish English, New Zealand English and Indian English. The factor extension turned out significant only in Indian English, while the factor polarity has an effect only in Jamaican English. Since it is exactly these two varieties that have a very small number of existentials with singular concord, as noted earlier, these results call for further research based on larger datasets.

Finally, the factor determiner type will be investigated. Following Hay and Schreier (2004), the present study distinguishes between seven different types of determiner, exemplified in (4.83): bare notional subjects, as in (a), and those preceded by a definite article, as in (b), a number, as in (c), a negative particle, as in (d), an adjective, as in (e), a quantifier beginning with *a* (including the indefinite article *a/an*), as in (f), or other quantifiers, as in (g).

(4.83) a. There's people here (ICE-GB:S1A-091)

b. Oh there are **the** boys getting out as well (ICE-GB:S1A-049)

c. There's **two** flats up over the pub (ICE-IRE:S1A-067)

d. There's **no** tanned men in them (ICE-GB:S1A-080)

e. And there's **big** inflatable dinosaurs (ICE-NZ:S1A-044)

f. There's **a lot of** deaths (ICE-IRE:S1A-036)

g. There's **so many** Germans (ICE-HK:S1A-011)

If there are several types of determiner before the notional subject, the one immediately following *there* + *be* determines the category the sentence is grouped in. The existential sentence in (d), for example, has a negative particle and an adjective before the notional subject but was categorized according to the first word *no*.

Determiner type has been investigated in many studies and in various varieties of English. An overview of the countries, the authors and their results is provided in Table 4.17. Note that it is difficult to compare these results because some of the classes of determiners contain different elements, that is, they are defined differently. Furthermore, the different orderings, especially those for the same variety of English, may also be due to "idiosyncrasies and sizes of the different data sets", as Hay and Schreier (2004: 232) suggest.

In the present study, determiner type taken as one single predictor did not emerge as significant in any of the samples. Looking at the individual types, however, significant results can be noted for Jamaican and Indian English. In the former variety, singular concord tends to be used more frequently when the notional subject is preceded by a negative particle. In Indian English, quantifiers with *a* tend to correlate with singular concord. These tendencies can also be observed when looking at the plots in Figure 4.20.[51] They represent the proportions of the seven determiner types

51 The corresponding token frequencies and percentages are given in Appendix 6.9.2.

in the different ICE samples, differentiated according to singular and plural concord. These plots show some further common effects. In line with previous studies (cf. Table 4.17), negatives and *a* quantifiers generally seem to favour singular rather than plural concord, although these factors did not emerge as significant in most of the samples. The latter factor seems to be particularly likely to co-occur with singular concord in the speech of the L2 English speakers.

Table 4.17: Effect of determiner type on singular concord (more > less) in different varieties of English (previous research).

CAN	*no* > number > other	Meechan/Foley (1994)
	a > definite > *no* > other	Walker (2007)
NZ	*a* > *no* > definite > number > bare > quantifier > adjective	Britain/Sudbury (2002)
	number > *no* > *a* > definite > bare > quantifier > adjective	Hay/Schreier (2004)
UK	partitive > *no* > definite > number > quantifier > bare	Tagliamonte (1998)

What is furthermore interesting to note is that quantifiers other than those beginning with *a* seem to occur frequently before notional subjects in all nine varieties and they tend to correlate with plural concord rather than singular concord. This finding also ties in with previous studies. As can be seen in Table 4.17, quantifiers tend not to have an effect on singular concord. This is very likely due to the fact that the quantifiers underline the plural meaning of the notional subject. In the present study, quantifiers are particularly frequent in the Hong Kong and the Indian samples. Recall that a similar preponderance of quantifying expressions in these varieties of English has been noted for left dislocation constructions (cf. section 4.1). It might be interesting to examine whether noun phrases are more frequently pre-modified by quantifiers in these varieties in general. In section 4.1, it has been suggested that the classifier system of Chinese might impact on Hong Kong English speakers' use of determiners and quantifying expressions. For Indian English, it has been noted that the quantifier *all* is used particularly frequently and has probably assumed new meanings and functions. Thorough investigations in this respect might be worth the effort.

Interestingly, numbers tend to co-occur with singular concord in some of the samples (Irish English, British English, Philippine English and Singapore English), although this factor did not turn out significant. This is exactly the type of determiner that emerged as most significant in Hay and Schreier's (2004) data from New Zealand English (cf. Table 4.83). What is surprising about these findings is that numbers also encode plural meaning, just like quantifiers. Possibly, the sequence '*there's* + number + plural notional subject' is developing into a fixed expression, at least for some speakers of English. In the present study, this is suggested by the Irish English data where most of the pre-modifying numbers occur in existentials with the contracted form *there's*.

In sum, the investigation of the determiner type in existentials with plural notional subjects yields some interesting results, although most of them do not emerge as statistically significant in the logistic regression model. In Jamaican English, negative

particles tend to covary with singular concord, which ties in with the finding that polarity has an effect on concord in this variety of English. Quantifiers with *a* emerge as significant only in Indian English, but they seem to favour singular concord also in the other varieties of English analyzed; especially so in the other L2 English varieties, which suggests that the word *a* at the beginning of the quantifier triggers singular meaning. Finally, numbers seem to often go with singular concord in Irish English, although the type does not emerge as significant.

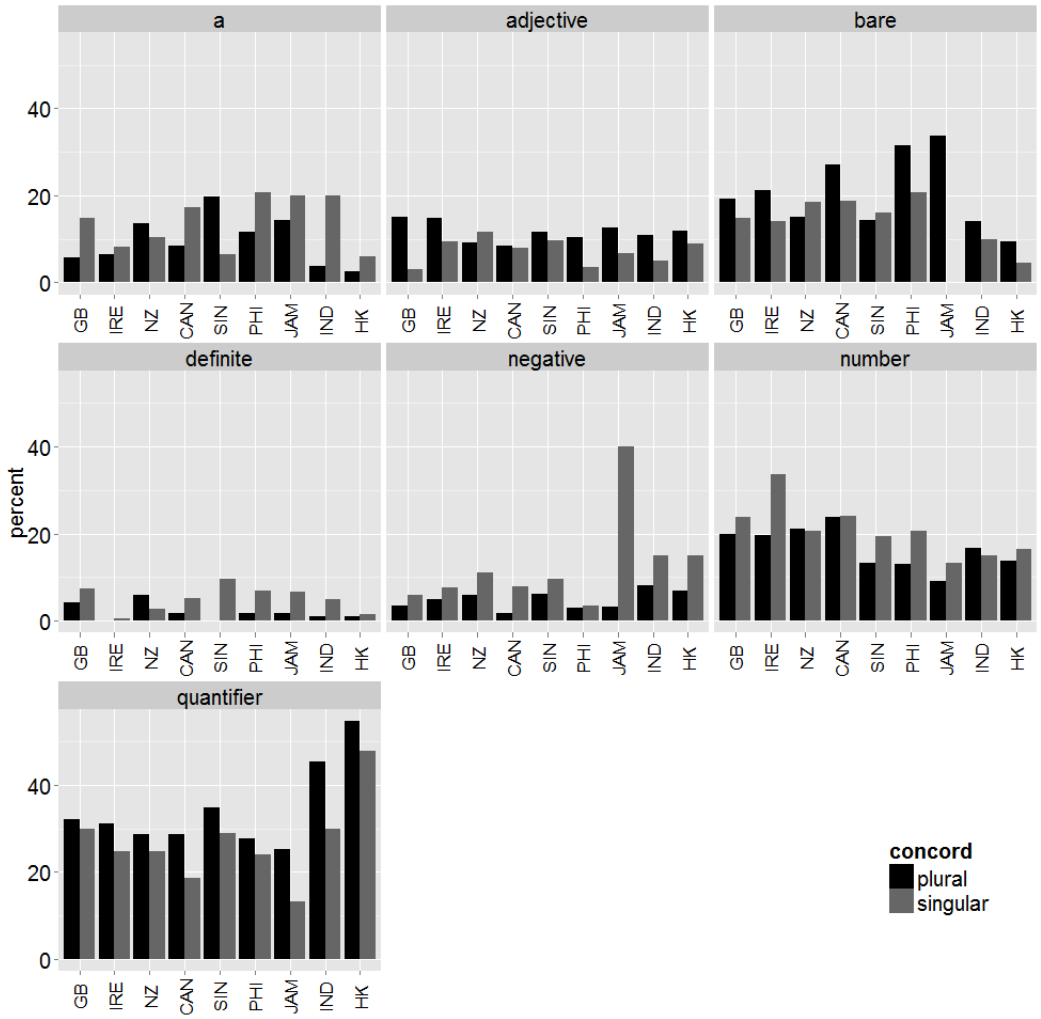

Figure 4.20: Effect of the factor determiner type on singular concord and plural concord.

4.4.4 Summary

The investigation of existential *there*-constructions in this section has shown that it is fairly frequent and more evenly distributed across the varieties of English analyzed

than the other information-packaging constructions that are examined in this study. The somewhat smaller frequencies attested for Singapore English and Jamaican English have been explained by the fact that speakers of these two varieties seem to use alternative constructions to express existence more frequently than do the speakers of the other varieties. In both cases L1 transfer seems to play an important role. In Singapore English, influence from the Chinese background language has made the verb *got* assume various new functions, including that of expressing existence. In Jamaican English, the influence form Jamaican Creole may be responsible for the more frequent use of *have* to express existence.

Differentiating between bare and extended existential clauses, it turns out that the L1 English speakers and, in particular, Jamaican English speakers show the highest proportions of the complex type. This finding is interesting in so far as similar patterns can be observed for left dislocation and fronting constructions, with the L1 varieties and Jamaican English showing the highest frequencies of complex structures.

Variation can also be found in terms of the preferred type of extension. Adverbial extensions and/or relative clause extensions are the most frequent types in all nine varieties of English. Jamaican English stands out among the other varieties of English analyzed in that it shows an extremely high proportion of relative clause extensions. For Irish English, it can be noted that participial extensions are relatively more frequent than in the other varieties. Since a similar preponderance can be attested for IT-clefts, it seems that Irish English speakers generally use *-ing* complementation patterns more frequently than do the other speakers (maybe together with New Zealand English speakers, who use more *ing*-forms in IT-clefts). An increasing frequency of *ing*-complements and a widening of the functional range of this complementation pattern is reported in the literature to apply to the English language generally (e.g. De Smet 2013; Duffley 2000; Fanego 1996, 2007; Mair 2002a, 2013; Rudanko 1998, 2000; among many others). The results of the present analysis suggest, however, that Irish English (and possibly New Zealand English) is more advanced in this development.

Among the existentials with relative clause extensions, special attention was given to subject relatives because for some varieties of English a preference for zero relative markers has been noted in the literature. The present study finds high proportions of zero-subject relatives in Irish English and Hong Kong English, which ties in with previous research. It is interesting to observe that those existentials are of very different kinds in these two varieties, however. While the sentences seem to be a learner feature in Hong Kong English, allowing the speakers to break down their sentences into smaller chunks and structuring them according to the principles of their topic-prominent L1, they seem to be compact constructions in Irish English. The suggestion of influence from Irish on the latter variety appears quite plausible (cf. Preusler 1938).

The investigation of variable concord in existentials with plural notional subjects yields a number of interesting results as well. Assuming that singular concord is on the rise in contemporary English (cf. Hay/Schreier 2004), it can be noted that the L1 English varieties, and in particular Irish, New Zealand and Canadian English, are more advanced in this development than the L2 English varieties analyzed. The comparatively low frequency of singular concord in the British English sample might be due to the somewhat higher level of formality of the 'direct conversation' files included in ICE-Great Britain (also see section 1.2 on data and methodology). Among

the factors that are reported in the literature to have an effect on singular concord, contractedness turns out to be most significant in the samples of the present study (e.g. *there's* vs. *there is*). The factor tense emerges as significant in Irish English and New Zealand English (with past tense favouring singular concord), exactly those two varieties which show the highest frequencies of singular concord. This suggests that singular concord is not restricted to cases of *there's* but occurs more variably, which in turn underlines the fact that these two varieties of English are further advanced in the development towards an increasing use of singular concord. The fact that tense emerges as significant in the Indian English sample as well is somewhat surprising. Further studies should test whether this finding is reliable or rather an artefact of the small sample size. The same holds for the factors extension and polarity, which seem to have an effect on concord in Indian English and Jamaican English, respectively, exactly those two varieties for which only very small numbers of existentials with singular concord can be attested.

The factor determiner type does not emerge as significant in any of the samples analyzed, but some tendencies can be observed when looking at the different types more closely. Negative particles and quantifiers with *a* (e.g. *a lot of*, *a couple of*) tend to covary with singular concord rather than plural concord; quantifiers without *a*, on the other hand, rather seem to go with plural concord. This suggests that quantifiers with *a* trigger a singular meaning because of the initial word *a*. It can be observed that quantifiers without *a* are particularly frequent in the Hong Kong English sample, which is interesting in so far as a similar pattern can be attested for left dislocation constructions, with quantifying LDs being more frequent in Hong Kong English than in the other varieties analyzed.

In sum, the data suggest that the expression *there's* is becoming more fixed, at least in the three L1 varieties Irish English, New Zealand English and Canadian English. This is in support of the claim that singular concord reflects lexicalization, a view that holds that the form *there's* has become a single lexical unit which is used in existentials regardless of the number of the notional subject (e.g. Hannay 1985; Meechan/Foley 1994; Schilling-Estes/Wolfram 1994; Crawford 2005). This development in turn may explain the large frequency of singular concord in these three varieties as compared to the other varieties analyzed. Particularly favourable environments for *there's* seem to be negative sentences, as in *There's not much jobs around for truckdriving is there* (ICE-NZ:S1A-012), and, in the case of Irish English, notional subjects preceded by a number, as in *There's two pools* (ICE-IRE:S1A-002).

4.5 Cleft constructions

This section examines in some detail three types of cleft construction: IT-clefts (e.g. *It's the writer <,> that uh gets you so involved*; ICE-GB:S1A-016), basic pseudo-clefts (e.g. *What I remember is we only had lunch*; ICE-PHI:S1A-053) and reversed pseudo-clefts (e.g. *That's what I think*; ICE-IND:S1A-004). The structure of the section is as

follows. First, a general overview of the distribution of these three cleft types across the varieties of English analyzed will be given, followed by a more detailed analysis of a number of syntactic features of each type of cleft.

4.5.1 Overall distribution

Table 4.18 and Figure 4.21 illustrate the distribution of IT-clefts, basic pseudo-clefts and reversed pseudo-clefts across the nine ICE samples analyzed.

Table 4.18: The distribution of IT-clefts, basic pseudo-clefts and reversed pseudo-clefts in the S1A-files of ICE (absolute token frequencies and frequencies per 100,000 words).

corpus	IT-clefts		basic PCs		reversed PCs	
	N	norm.	N	norm.	N	norm.
GB	60	29.8	124	61.5	244	121.0
IRE	109	54.0	71	35.2	270	133.9
NZ	53	25.2	95	45.1	300	142.4
CAN	48	20.9	126	54.8	297	129.1
SIN	31	15.2	97	47.7	223	109.7
PHI	50	23.1	133	61.4	307	141.7
JAM	72	33.8	163	76.4	296	138.8
IND	30	13.9	141	65.2	207	95.8
HK	24	10.1	54	22.7	116	48.7

Turning to reversed pseudo-clefts first, it can be observed that they are the by far most common type of cleft construction in all corpora analyzed. Except for Indian English, Singapore English and Hong Kong English, the frequencies per 100,000 words amount to over 120 tokens in all varieties. New Zealand English shows the highest frequency with 142.4 tokens per 100,000 words. According to chi-squared tests, the frequency of reversed pseudo-clefts in New Zealand English is significantly different only from the frequencies in Hong Kong English (at the level $p < 0.001$), Indian English ($p < 0.01$) and Singapore English ($p < 0.05$).

Basic pseudo-clefts are far less common than the reversed type. Jamaican English shows the highest frequency (76.4 tokens per 100,000 words), followed by Indian English (65.2), British English (61.5), Philippine English (61.4) and Canadian English (54.8). The differences between these varieties of English are not statistically significant. Yet, the frequencies of Jamaican English and the other varieties are significantly different (New Zealand and Singapore English at $p < 0.01$; Irish and Hong Kong English even at $p < 0.001$).

Turning to IT-clefts, it is interesting to note that this type of cleft is far more frequent in Irish English than in the other varieties analyzed and that it is the only variety where we find more IT-clefts than basic pseudo-clefts (54.0 vs. 35.2 tokens per 100,000 words). The frequencies of IT-clefts in the other varieties are significantly

lower than in Irish English (Jamaican English and British English at the level $p < 0.05$; New Zealand English at $p < 0.01$; and all other varieties at $p < 0.001$). The high frequency of IT-clefts in Irish English does not come as a surprise as the widespread use of clefting has repeatedly been reported in the literature for this variety of English and other Celtic Englishes (e.g. Visser 1963; Filppula 1999, 2006, 2009, 2012; Hickey 2007; Filppula/Klemola 2012). More on this below in the subsection on IT-clefts.

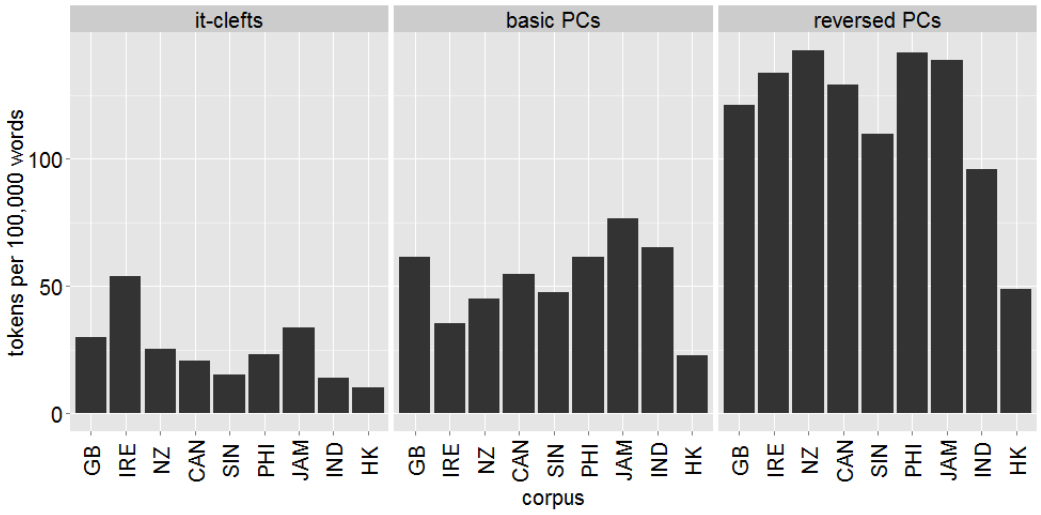

Figure 4.21: IT-clefts, basic pseudo-clefts and reversed pseudo-clefts in the S1A-files of nine ICE corpora (frequencies per 100,000 words).

Hong Kong English stands out among the varieties of English analyzed in that it shows very low frequencies of all three types of cleft construction. Singapore English also shows relatively low frequencies of clefts. This might, at least in part, be due to influence from the Chinese L1. For one, relative clauses in Chinese are structured very differently from those in English. While relative clauses are postmodifying in English, they are premodifying in Chinese, realized with a nominalizer that comes before the head noun. Sentence (4.84) exemplifies a Cantonese relative clause, the one in (4.85) is from Mandarin.

(4.84) *sīk* *Gwóngdūng-wá* *ge* *hohksāang* *háau* *dāk* *hóu* *dī*
know Cantonese that students examine ADV well a-bit
'The students who know Cantonese did better (on the exam).'

(adapted from Matthews/Yip 1994: 326)

(4.85) *zhòng* *shuǐguǒ* *de* *nóngrén*
grow fruit NOM farmer
'(the) farmer(s) who grow fruit'

(adapted from Li/Thompson 1981: 580)

In (4.84), the nominalizer *ge* comes before the head noun *hohksāang* 'students'. The same is the case in the Mandarin sentence: the nominalizer *de* comes before the head noun *nóngrén* 'farmer'. Because of this difference "many Chinese ESL learners in Hong Kong have problems with complex English sentences involving relative clauses" (Chan 2004: 59) and this might make them avoid cleft structures.

For the low frequency of IT-clefts there might be additional reasons. First, Chinese is a pronoun-dropping language and does not syntactically require dummy subjects. Hence, the *it* in IT-clefts has no equivalent in Chinese (cf. Li/Thompson 1981: 91; Huang 1984). Second, in English IT-clefts, the focused element is placed after *it be*. The Cantonese *haih…ge* and the Mandarin *shì…de* structures, on the other hand, involve no such word order change. Consider the following sentences for illustration, a Cantonese *haih…ge* sentence in (4.86) and a Mandarin *shì…de* sentence in (4.87).

(4.86) *ngóh haih hohk Wihngchēun ge*
I am learn Wing Chun SFP
'It's Wing Chun I learn.'

(Matthews/Yip 1994: 356)

(4.87) *Zhāngsān shì zuótiān lái de*
Zhangsan COP yesterday come DE
'It was yesterday that Zhangsan came.'

(Hole 2011: 1707)

In these sentences, the focused elements are *Wihngchēun* 'Wing Chun' and *zuótiān* 'yesterday', respectively. In the English sentences, these words are therefore placed after *it was*. Compare the word order in the canonical counterparts: *I learn Wing Chun* and *Zhangsan came yesterday*. Note that there is no such word order change in the Chinese sentences. Here, the copular verbs *shì* and *haih* are simply inserted before the focused elements. This challenge of moving a constituent to the periphery of the clause, coupled with the lack of dummy subject *it*, might be the reason why only a relatively small number of IT-clefts can be attested for Singapore English and Hong Kong English. Because of the different proficiency levels of the speakers in these two countries the influence from the Chinese language is particularly strong on Hong Kong English.

Indian English also shows fairly low numbers of IT-clefts. This may be accounted for by the fact that the restrictions on word order are much more relaxed in Indian languages which, in turn, may have an influence on word order in Indian English (clefts typically occur in languages with more fixed word order; cf. e.g. Lambrecht 2001b). Support for this hypothesis also comes from the fact that high frequencies of left dislocation and, in particular, fronting constructions can be attested for this variety of English. It seems that speakers of Indian English prefer these types of constructions over IT-clefts to focus on clause constituents.

After this general overview of the distribution of cleft constructions across the nine varieties of English analyzed, the three types of cleft will be discussed separately and in some more detail in the following subsections.

4.5.2 It-clefts

As noted above, the most striking finding with respect to IT-clefts is their overwhelmingly higher incidence in Irish English than in the other varieties analyzed. Influence from the Irish language is commonly given as an explanation for the high incidence and flexible usage of IT-clefts in Irish English. Filppula (2012) and Filppula and Klemola (2012) even argue that the development of the IT-cleft construction in English generally is due to contact with Celtic languages. Evidence in support of this hypothesis comes from the fact that the cleft construction - or its Celtic equivalent - had already existed in Celtic languages long before it was first attested in the English language (Filppula/Klemola 2012: 1696). Another factor speaking for Celtic influence is the "prominence of clefting in present-day and earlier Celtic-influenced varieties of English" (Filppula/Klemola 2012: 1698). A third factor Filppula and Klemola give concerns the structural and functional similarity of the Celtic cleft and the English cleft. For illustration consider the sentences in (4.88) (cf. example (3.70) in section 3.4). The Irish cleft is a copula construction that was developed in early Irish for the purpose of focussing certain constituents in the clause.

(4.88) *is* *í* *mo* *dheirfiúr* *a* *chonaíonns* *i* *Sasana*
COP her my sister REL live-REL in England
'It's my sister that lives in England.'

(Stenson 1981: 99)

Irish copula constructions are more flexible with regard to the type of clefted element they allow. That is, basically any constituent but finite verbs can be clefted (Stenson 1981). This made researchers argue that it is not so much the frequency of IT-clefts as such that can be traced back to Irish influence but rather the realizational variation, and that it is this wider scope of realizational options that is unique to Irish English. Hickey (2007), for example, notes that "[t]he range of clefting options in Irish is large indeed and it may well be that it is the scope, rather than just the fact of clefting, which is attributable to Irish" (2007: 269). Filppula and Klemola (2012) argue in a similar way:

> Of course, this construction [i.e. IT-cleft; cw] is, and has for a long time, been part of Standard English (StE) grammar, but what gives Irish English and some of the other Celtic Englishes a distinctive flavor is that the uses of the *it*-cleft construction are both functionally and syntactically less restricted than in StE, in particular. In this respect, they behave much like their Celtic counterparts, which is a major factor speaking for Celtic substratum influence. (Filppula/Klemola 2012: 1698)

Beal (2012) agrees that some features of IT-clefts must be explained by substrate influence. In her study, it is the choice of subordinator that she identifies "as a likely candidate for substratal Irish influence" (2012: 175).[52]

52 Beal's (2012) study is also based on ICE, but she reports surprisingly different frequencies of IT-clefts. She finds 69 instances of IT-clefts in the private dialogues files of ICE-Great Britain, 67 tokens in ICE-India, 99 tokens in ICE-Ireland, 95 items in ICE-Jamaica and 33 tokens in ICE-Singapore (2012: 163). The discrepancy between some of these numbers and those of the present study is staggering. Probably, the studies are based on slightly different definitions of the IT-cleft construction. Beal might

Adding to Beal's (2012) study, a number of linguistic features of IT-clefts will be examined in the following subsections. These features include the syntactic category and the syntactic function of the clefted element and the choice of subordinator. Furthermore, the sentence type in which IT-clefts occur will be examined (declarative vs. interrogative sentence). The question of whether Irish English shows a wider scope of clefting options than other varieties of English is one of the guiding questions in the following investigations.

The syntactic category of the clefted element

Table 4.19 shows the distribution of IT-clefts according to the syntactic category of the clefted element. These include noun phrases with a proper noun, as in (4.89a), or a pronoun, as in (b) and (c), prepositional phrases, as in (d), adjective phrases, as in (e), and adverb phrases, as in (f). The latter two categories occur so infrequently in the data that they are grouped together as APs. Furthermore, there are clefted clauses, as in (g), but they are also very infrequent in the present data.

(4.89) a. It is **the society** that controls (ICE-IND:S1A-011)

b. It was **you** that told me that (ICE-GB:S1A-099)

c. **What** is it that you do on campus (ICE-JAM:S1A-045)

d. It was **in September** we used to go (ICE-IRE:S1A-088)

e. It was **fucking dodgy wicked** he got (ICE-NZ:S1A-047)

f. It's only **now** that you're realising a lot of other little things (ICE-NZ:S1A-046)

g. It's **only because they knew him in the bank** why they did it (ICE-JAM:S1A-013)

The figures in Table 4.19 indicate that Irish English speakers do not behave remarkably differently from the speakers of the other varieties. Clefted nouns are generally most common, accounting for more than half of all IT-clefts in all varieties but Jamaican English, where we find only 48.6% of clefted nouns. In Hong Kong English, on the other hand, they even make up nearly 90% of all IT-clefts. This preponderance of nominal clefted elements ties in with previous studies and grammar books which also identify them as the most frequent type of clefted element (e.g. Ward et al. 2002).

There are some more highlighted clauses in the Irish English data than in most other data, but the frequency is very low and there are just as many clauses in Jamaican English. The latter variety contains the highest proportion of pronoun IT-clefts (36.1% of

have also included inferential clefts in her counts, as in *It's just if you're doing phonetics analysis you know you should never record in Dolby* (ICE-GB:S1A-008; for more information on inferential clefts see, for example, Delahunty 1995). Despite these different frequencies we can still observe some common tendencies: among the varieties of English analyzed by both studies, Irish English shows the highest frequency of IT-clefts, followed by Jamaican English, then British English and finally Indian English and Singapore English.

all IT-clefts), with most of these containing an interrogative pronoun, as exemplified in sentence (c) above (18 items or 69.2% out of all pronominal IT-clefts). Interrogative clefts will be discussed in some more detail below.

Adjective and adverbial phrases (AP) and prepositional phrases (PP) are rather infrequent in all corpora, with the proportions of APs being highest in Philippine English (12.2%) and the proportions of PPs in Indian English (10.3%). Note, however, that these percentages correspond to only six and three items, respectively.

Table 4.19: The syntactic category of the clefted constituent in IT-clefts (token frequencies and percentages out of all IT-clefts).

corpus	noun		pronoun		PP		AP		clause	
	N	%	N	%	N	%	N	%	N	%
GB	40	66.7	13	21.7	3	5.0	3	5.0	1	1.7
IRE	80	73.4	14	12.8	5	4.6	5	4.6	5	4.6
NZ	36	67.9	13	24.5	2	3.8	2	3.8	-	-
CAN	38	79.2	6	12.5	1	2.1	1	2.1	1	2.1
SIN	20	64.5	5	16.1	2	6.5	2	6.5	2	6.5
PHI	34	68.0	8	16.0	1	2.0	6	12.0	1	2.0
JAM	35	48.6	26	36.1	5	6.9	1	1.4	5	6.9
IND	17	56.7	7	23.3	3	10.0	1	3.3	2	6.7
HK	21	87.5	1	4.2	-	-	2	8.3	-	-

In sum, the analysis of the syntactic category of the clefted constituent in IT-clefts cannot confirm the claim that Irish English allows for a wider range of elements to be clefted (e.g. Bliss 1979; Quirk et al. 1985; Filppula 1999; Hickey 2007; Filppula/ Klemola 2012). It may well be that this is only true for vernacular Irish English. That is, the small number of adjective phrase clefts and the absence of verbal clefts (e.g. *I think it was painting I was*; Filppula 1999: 260) may be due to the fact that the ICE corpus contains data of educated speakers of standard English. Even the face-to-face conversations seem not to be informal enough to get a larger number of such items (also see Siemund and Beal (2011: 258) on the (near-)absence of these types of cleft in their data). Rather, it is Jamaican English that shows some more variation than the other varieties analyzed. It has the lowest proportion of nominal clefted elements but a substantial number of pronominal items.

The syntactic function of the clefted element

The analysis of the syntactic function of the clefted element yields the results presented in Table 4.20. The syntactic functions annotated in the present study comprise subjects, as in (4.90a), objects, as in (b), adverbials, as in (c), and complements, as in (d).

(4.90) a. It was **you** opened the curtains (ICE-IRE:S1A-050)

b. It's **Word Word Five** I have (ICE-IRE:S1A-047)

c. It was **during first year** I saw him (ICE-SIN:S1A-082)

d. So it's **crazy** they're you know (ICE-JAM:S1A-014)

According to Quirk et al. (1985: 1385), the most common types of clefted constituents include subjects, direct objects, as well as time and place adverbials. This is corroborated by the findings of the present study: only three clefted complements have been found in the nine samples, all other clefted elements are either subjects, objects or adverbials. As can be seen in Table 4.20, subject clefts make up the highest proportion in all corpora. In Hong Kong English and Philippine English they account for nearly 60% of all IT-clefts. Irish English and Jamaican English, on the other hand, show the smallest proportions with only around 38% of subject clefts.

Table 4.20: The syntactic function of the clefted constituent in IT-clefts (token frequencies and percentages out of all IT-clefts).

corpus	subject		object		adverbial		comp.	
	N	%	N	%	N	%	N	%
GB	30	50.0	14	23.3	16	26.7	-	-
IRE	41	37.6	33	30.3	35	32.1	-	-
NZ	23	43.4	12	22.6	17	32.1	1	1.9
CAN	24	50.0	10	20.8	13	27.1	1	2.1
SIN	15	48.4	3	9.7	13	41.9	-	-
PHI	29	58.0	4	8.0	17	34.0	-	-
JAM	27	37.5	26	36.1	18	25.0	1	1.4
IND	13	43.3	4	13.3	13	43.3	-	-
HK	14	58.3	5	20.8	5	20.8	-	-

Object clefts occur remarkably infrequently in the L2 English varieties analyzed, with the exception of Jamaican English, which in fact shows the highest proportion of object clefts among the nine varieties. The lack of object clefts might be due to processing reasons. Comparing subject and object relative clauses, studies have found that in SVO languages object extraction causes increased processing difficulty because it carries a longer dependency (Warren/Gibson 2005). This might explain the general preference of subject clefted elements over object clefted elements, and in particular in the L2 varieties. As noted earlier, L2 English speakers might shy away from using cleft constructions generally because of their complexity. If it was the case that object clefts lead to even more processing load than subject clefts, it would not at all be surprising to find less object clefts in the speech of L2 English speakers than in that of L1 speakers. In the case of Singapore English, the preference of subject and adverbial clefts over object clefts may have an additional reason. It may well be due to influence from the Mandarin substrate because the Mandarin *shì…de* cleft construction freely allows subjects and adverbials to be clefted but not objects (Li 2008: 763). Objects can also be clefted, but this requires some deviations from the basic word order.

What is interesting to note about Irish English is that the proportions of subject,

object and adverbial clefts are most balanced among the varieties analyzed. That is, the percentages are closest together, which means that the three types of syntactic function occur to roughly the same extent. A similarly balanced usage pattern can be attested for Jamaican English. Beal (2012: 165) regards the similarity between these two varieties of English as a result of Irish and Irish English contact with the developing Jamaican English, a plausible explanation considering that other features of Caribbean English have already been shown to be closely linked with Irish and Irish English features (e.g. Rickford 1986).

Regarding the question of whether Irish English shows a wider scope of clefting options, then, we can say that Irish English indeed stands out in that it shows more variation or a more balanced usage of subject, object and adverbial clefts.

Subordinator type

Another property of IT-clefts worth analyzing concerns the choice of subordinator.[53] The major types of subordinator that occur in IT-clefts include *wh*-words (e.g. *who*, *which*, *when*), *that*, zero, and *ing*-forms. The different types are illustrated in (4.91) below. The use of *ing*-forms is, of course, also a case of zero marking. Those tokens are nevertheless counted separately because they constitute particularly interesting exemplars from a cross-varietal perspective. While an increasing use of *ing*-complements and a widening of its functions has been reported to be a general trend since Early Modern English (e.g. De Smet 2013; Duffley 2000; Fanego 1996, 2007; Mair 2002a, 2013; Rudanko 1998, 2000; among many others), the examination of extended *there*-existentials in the present study suggests that Irish English is somewhat more advanced in this development (cf. section 4.4). The analysis of a specific type of pseudo-cleft of the form *What they do is (to) travel around the world*, on the other hand, reveals a slightly higher proportion of *ing*-complements in New Zealand English (Mair/Winkle 2012). Hence, it is interesting to investigate in how far *ing*-complements are used in IT-clefts.

(4.91) a. It was actually me **who** picked it. (ICE-NZ:S1A-001)

b. It's not me **that**'s got to act. (ICE-GB:S1A-062)

c. It is the government **realises** he's really dangerous. (ICE-GB:S1A-049)

d. It's like an heart attack **he had** . (ICE-JAM:S1A-074)

e. It was actually him **trying** to ring. (ICE-NZ:S1A-091)

Within the category 'zero', subject clefts are to be distinguished from object clefts (cf. (c) vs. (d)). According to Quirk et al. (1985: 1250), the former type is "of doubtful acceptability". In a similar way, this type has been considered completely intolerable by native speaker informants in Siemund and Beal's (2011: 255) study, yet still quite

53 The cleft clause superficially looks like a restrictive relative clause. Yet, since it is debatable whether it really is a relative clause or not (cf. section 3.3), I decided to refer to the 'relative marker' as subordinator.

a number of such items could be attested for their Irish English data (18 of the 57 subject clefts have zero subordinators).

The distribution of the different subordinator types across the varieties of English analyzed are presented in Figure 4.22.[54] The category 'other' comprises incomplete constructions (i.e. those where the cleft clause is absent) and clefts with the subordinators *since*, *till*, *for* and *and*, which are rarely found in the present data. As can be seen, there is quite some variation of subordinator choice across the nine varieties of English analyzed. In all corpora but ICE-Ireland overt marking with *that* or *wh*-words is most common, making up more than half of all instances. In the Irish component, on the other hand, zero marking is the by far most frequent variant. More on this variant below.

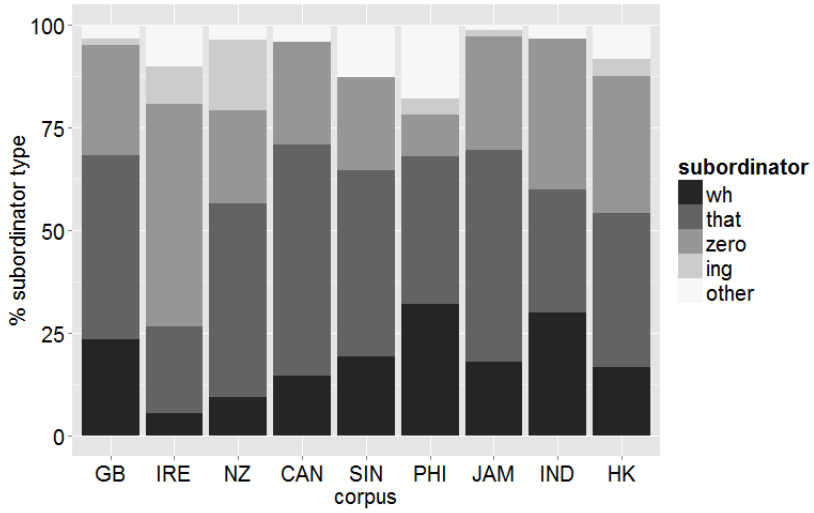

Figure 4.22: Different types of subordinator in IT-clefts (percentages out of all IT-clefts).

What is furthermore interesting to note is that *ing*-forms occur most frequently in New Zealand English (17.0% of all IT-clefts), followed by Irish English (9.2%). The proportions of this variant are far smaller in all other varieties, ranging from 0% to 4%. Recall that in the case of existential *there*-constructions it was Irish English that showed the highest proportion of participial extensions. Given these findings, it might be said that New Zealand English and Irish English are more advanced in the development towards an increasing use of *ing*-complements.

The fairly high proportion of items of the category 'other' in Philippine English is mainly due to incomplete or truncated IT-clefts, of which speakers of this variety seem to make more use than the speakers of the other varieties.

Coming back to zero subordinators, the distribution of this subordinator type in subject clefts is presented in Table 4.21. The three rows give the token frequencies of subject IT-clefts, the token frequencies of subject IT-clefts with zero subordinator and their percentages out of all subject IT-clefts.

54 The token frequencies of the different subordinator types and their proportions within each ICE sample can be found in Appendix 6.10.3.

Table 4.21: The frequency of subject IT-clefts with zero subordinator (token frequency of subject IT-clefts, frequency of subject IT-clefts with zero subordinator and their percentage out of all subject IT-clefts).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
subject clefts	30	41	24	23	15	28	27	16	14
zero tokens	4	12	2	4	2	1	5	4	6
% of zero	13.3	29.3	8.3	17.4	13.3	3.6	18.5	25.0	42.9

Note, first of all, that in all corpora the token frequencies of subject clefts with zero subordinator are very small, ranging from only 1 to 12 items. Furthermore, subject clefts in general are infrequent in the Singapore English, Indian English and Hong Kong English data. Hence, reliable conclusions are difficult to be drawn. What can definitely be said, however, is that among the L1 varieties the Irish component is the outlier, with nearly one third of all subject IT-clefts having zero subordinators (12 items or 29.3%). Surprisingly, the proportion of this type of subject cleft is nearly as high in Indian English (25.0%) and even higher in Hong Kong English (42.9%). Given these findings it is difficult to straightforwardly agree with Beal (2012), who identifies subordinator choice as a likely candidate for substratal Irish influence. The picture is more complex and several plausible explanations are possible. For one, the occurrence of zero subject clefts could be described as a learner feature (simplification), in line with Siemund and Beal (2011). Second, substratal influence may be responsible for the somewhat higher frequency in some varieties, despite the completely different background languages (in line with Beal (2012)). The high incidence of zero subordinators in Irish English can plausibly be traced back to Irish, as the following examples from Preusler (1938) illustrate.

(4.92) *John sydd yn gryf*
'[it is] John [who] is strong'

(4.93) *pwy sydd yma?*
'who [is it that] is there?'

(Preusler 1938: 184)

As noted earlier, Irish uses such copula constructions to highlight certain elements of the clause. Note that the relative markers are omitted in the Irish sentences; this is also freely permitted in subject relatives. According to Preusler (1938: 185), English contact clauses - that is, relative clauses without relative markers - can be traced back to Irish influence.

Influence from the background language may also have impacted on relatives in Hong Kong English. Hung (2012) notes that relative clauses without relative pronouns are "common even among highly educated speakers of HKE" (2012: 127). He suggests that this may be due to Chinese influence because in Chinese, relative clauses have no relative pronouns. Similarly, Newbrook (1988, 1998) finds zero relatives to be a frequent feature in Hong Kong students' writings. Hence, it may well be that language contact has led to the same outcome in both varieties Irish English and Hong Kong English despite the different contact languages: a somewhat higher frequency

of zero subject clefts. Of course, it may also be a combination of both influence from the substrate and the outcome of language acquisition that impacts on the somewhat higher incidence of zero subject clefts.

Sentence type

Finally, some variation with respect to sentence type will be discussed. Typically, IT-clefts occur in declarative sentences, but in all corpora there are also some interrogative items. Interrogative clefts are usually introduced by the question words *what*, *why* or *who*, rarely also *where*. The subordinator is usually *that*, sometimes also zero. Examples are given in (4.94).

(4.94) a. **What** is it that you love about Saint Mary? (ICE-JAM:S1A-057)

b. But **why** is it always walnut cake we order? (ICE-SIN:S1A-006)

c. **Who** was it that came up for my birthday in February? (ICE-CAN:S1A-028)

d. **Where** was it we were and we saw [two words unclear]? (ICE-IRE:S1A-077)

Siemund and Beal (2011) note that "[q]uestion-word clefts are very common, and nowhere more so than in standard forms of spoken English" (2011: 254). The findings of the present study confirm this claim because the proportions of interrogative clefts out of all IT-clefts are much higher in the present analysis - which is based only on spoken data - than those reported by Siemund and Beal - who consider all files of the ICE corpora, that is, speech and writing. The percentages of interrogative clefts in their data are as follows: 5.32% of all IT-clefts in ICE-Great Britain, 3.31% in ICE-Ireland and 0.68% in ICE-India. The results of the present study are given in Table 4.22. As can be seen, the percentages of interrogative clefts are much higher than those reported by Siemund and Beal. They are particularly common in Jamaican English, accounting for 23.6% of all IT-clefts.

Table 4.22: Interrogative IT-clefts in the S1A-files of nine ICE corpora (absolute token frequencies and percentages out of all IT-clefts).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
tokens	11	14	7	3	6	6	17	4	3
%	18.3	12.8	13.0	6.3	19.4	12.0	23.6	13.3	12.5

Interestingly, the questions in the Hong Kong English data and most of the questions in the British English data have no introductory *wh*-word but begin with a form of *be*. Consider the sentences in (4.95) for illustration.

(4.95) a. Was it you just ringing up? (ICE-GB:S1A-078)

b. Is it now five months for Janet pregnant? (ICE-HK:S1A-014)

c. Is it this term her first year? (ICE-HK:S1A-023)

d. Is it ninety-one you are here right? (ICE-HK:S1A-024)

Admittedly, sentences (b) and (c), taken from ICE-Hong Kong, are rather debatable examples of IT-clefts because the cleft clauses are no proper clauses. However, it was decided that they sufficiently resemble clefts proper to be included in the present study.

Summary

Summing up the investigation of IT-clefts, the most striking finding to note is the overwhelmingly higher frequency in Irish English than in the other varieties of English analyzed. While subject clefts are generally preferred over object clefts and adverbial clefts, Irish English speakers show some more variation than the other speakers in that they use these three types of cleft to roughly the same extent. Irish English also stands out in terms of subordinator choice. It shows the largest number of zero subordinators in both object and subject position, which is very likely due to influence from Irish.

Jamaican English, the variety with the second highest frequency of IT-clefts, also shows some variation with regard to the syntactic function and syntactic category of the clefted element and it seems to have a number of features in common with Irish English, which may be the result of contact with Irish and Irish English when Jamaican English was just beginning to develop. Jamaican English also shows the highest number of interrogative clefts, which further underlines the impression that IT-cleft constructions are flexibly used in this variety of English. Regarding IT-cleft frequency and its variable usage, Jamaican English is clearly the outlier among the L2 varieties analyzed. The construction is particularly infrequent in Indian English, Singapore English and Hong Kong English. This might be due to processing reasons because cleft sentences are rather complex structures and might therefore be dispreferred by learners of English. Furthermore, the low frequencies may be accounted for by influence from the background languages.

4.5.3 Basic pseudo-clefts

As noted earlier, basic pseudo-clefts are more frequent than IT-clefts in all varieties of English analyzed but Irish English. They are particularly frequent in Jamaican English, but the difference in frequency in comparison to the other varieties is not as remarkable as the one that can be attested for Irish English in the case of IT-clefts. The following subsections will examine a number of linguistic features of the construction in order to reveal similarities and differences across the varieties analyzed.[55]

55 The discussion of pseudo-clefts is relatively brief, but I have thoroughly examined them elsewhere; see Mair and Winkle (2012) and Winkle (2011).

Types of basic pseudo-cleft

The present study examines different types of basic pseudo-cleft, categorized according to the initial elements. These include pseudo-clefts introduced by *wh*-words (e.g. *what*, *how*), *all* and pro-nouns (e.g. *thing*, *person*, *one*, *reason* etc.), accordingly also referred to as WH-clefts, ALL-clefts and TH-clefts. The three types are illustrated in (4.96).

(4.96) a. **What** I really liked were the dresses (ICE-PHI:S1A-016)

b. **All** you've to do is look them up (ICE-IRE:S1A-084)

c. **The only thing** I worry about is my health (ICE-CAN:S1A-015)

d. **The person** who gained mainly was the agent (ICE-JAM:S1A-088)

The variant with *all* has only a slightly different meaning than the equivalent with *what*. Compare sentence (b) with *What you've to do is look them up*. In TH-clefts, we have a noun phrase with a pro-noun as head (e.g. *person, one, place, time, reason, way*) and a relative clause as modifier. Sometimes there are corresponding WH-clefts which are very similar in meaning. Compare sentence (c), for example, with *What I worry about is my health*. Yet, the pro-nouns in TH-clefts often contain some additional meaning not expressed by *what*, such as *only* in (c), or *the best/first/other thing*. In sentence (d), the equivalent with *who* is not possible at all because *who* is not admitted in fused relatives: **Who gained mainly was the agent*. The distribution of these different types of basic pseudo-cleft across the nine varieties of English analyzed is illustrated in Figure 4.23.[56]

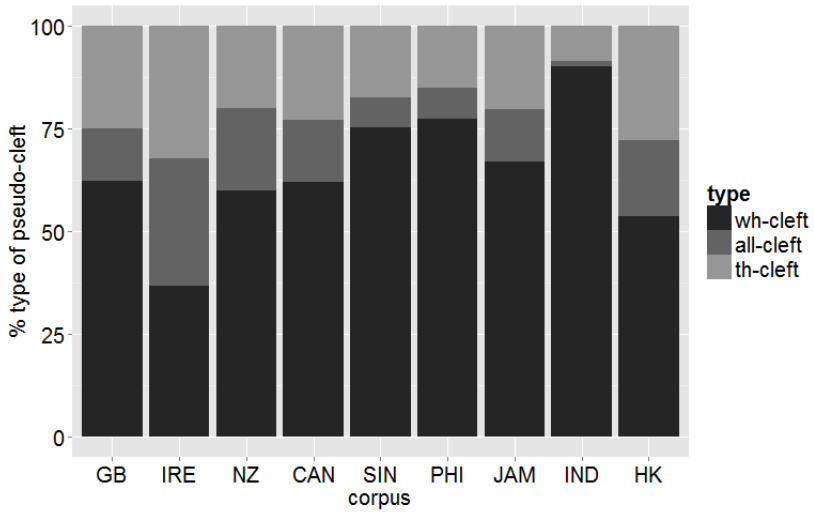

Figure 4.23: The proportions of basic pseudo-clefts introduced by *wh*-words, *all* and pro-nouns (percentages out of all basic pseudo-clefts).

56 The token frequencies of the different types of basic pseudo-cleft and their proportions within the ICE samples are given in Appendix 6.10.3.

It has been noted before that Irish English deviates from the other varieties analyzed in that it is the only variety for which more IT-clefts than basic pseudo-clefts can be attested (54.0 vs. 35.2 tokens per 100,000 words, respectively). Irish English also stands out among the other varieties in that it shows the greatest variation as far as pseudo-cleft type is concerned, as can be seen in Figure 4.23. In all the other varieties, basic pseudo-clefts with introductory *wh*-words are the most commonly used variant, accounting for more than half of all basic pseudo-clefts. The preponderance of WH-clefts is particularly dominant in Indian English, where they make up more than 90% of all basic pseudo-clefts. In Irish English, on the other hand, we find more variation in the sense that ALL-clefts and TH-clefts account for larger proportions than in the other varieties analyzed. Interestingly, Irish English even has the highest absolute token frequency of ALL-clefts (22 tokens or 31.0% of all basic pseudo-clefts), followed by Jamaican English with 21 tokens (12.9%) and New Zealand English with 19 tokens (20.0%). Note that the reversed type of ALL-cleft can also most frequently be attested for the Irish English data. This suggests that the (basic and reversed) ALL-cleft is a much more well-established variant of pseudo-cleft in Irish English than in the other varieties of English analyzed.

As for TH-clefts, they account for nearly one third of all basic pseudo-clefts in Irish English, while the proportions in the other varieties are all below 30%. The great majority of the TH-cleft tokens in Irish English - and in all other varieties - contain the pro-nouns *thing* or *reason*. In sum, it is quite surprising to find so much variation in the use of basic pseudo-clefts in Irish English because the overall number of this type of cleft is relatively small in this variety of English.

Regarding Indian English, the extremely high proportion of WH-clefts is remarkable, as has already been pointed out above. It seems that in Indian English WH-clefts are used as rather formulaic expressions. I got this impression while examining the verbs in the cleft clauses of WH-clefts. The results of this analysis will be presented in the following subsection.

Verb types in wh-clefts

This section concentrates on WH-clefts only. The investigation of the verbs in the cleft clauses yielded six major semantic groups, namely 'do', 'happen', 'talk', 'opinion', 'feel' and 'realize'. The two former categories only contain the verbs *do* and *happen*, respectively; the category 'talk' includes verbs of speaking such as *say*, *talk*, *tell*, *ask*; the category 'opinion' includes verbs with which the speakers express their opinion or want to make a point, for example, *think*, *mean* and *believe*; the category 'feel' includes verbs and adjectives which express negative and positive feelings, such as *love*, *like*, *hate* or *what is annoying/interesting/strange*; the category 'realize' includes verbs with which the speakers express that they have realized, learned or understood something, for example, *realize*, *see*, *hear*, *find*, *learn*, *discover*. The distribution of WH-clefts according to these semantic categories is presented in Table 4.23 (absolute frequencies and percentages out of all WH-clefts).

In all varieties but Jamaican and Indian English, WH-clefts with the verb *do* in the

cleft clause make up the highest proportion. They are particularly dominant in Canadian English and New Zealand English, accounting for more than 50% of all WH-clefts in the former variety and for more than 42% in the latter.[57] In Jamaican English the *do*-type is also quite common (22.9%). It is only topped by the category 'other' (26.6%), which suggests that there is more variation with respect to verb type in Jamaican English than in the other varieties.

Table 4.23: WH-clefts according to the type of verb in the cleft clause (absolute frequencies and percentages out of all WH-clefts).

corpus*	do		happen		talk		opinion		feel		realize		other	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
GB (77)	25	32.5	11	14.3	8	10.4	4	5.2	10	13.0	1	1.3	18	23.4
IRE (26)	8	30.8	5	19.2	4	15.4	1	3.8	6	23.1	-	-	2	7.7
NZ (57)	24	42.1	14	24.6	4	7.0	3	5.3	5	8.8	1	1.8	6	10.5
CAN (78)	39	50.6	9	11.7	4	5.2	3	3.9	6	7.8	3	3.9	14	18.2
SIN (73)	23	31.5	13	17.8	6	8.2	12	16.4	6	8.2	-	-	13	17.8
PHI (103)	33	32.0	7	6.8	10	9.7	13	12.6	16	15.5	5	4.9	19	18.4
JAM (109)	25	22.9	14	12.8	17	15.6	6	5.5	1	0.9	17	15.6	29	26.6
IND (127)	20	15.9	30	23.8	23	18.3	28	22.2	2	1.6	11	8.7	13	10.3
HK (29)	11	37.9	2	6.9	3	10.3	4	13.8	1	3.4	-	-	8	27.6

* The numbers in brackets after the corpus labels give the frequencies of WH-clefts in each corpus.

In Indian English, we find high proportions of the *happen*-type (23.8%) and of verbs of the categories 'opinion' (22.2%) and 'talk' (18.3%). The latter two categories are basically represented by four verbs only: *think*, *mean*, *feel* and *say*. The verbs *think* and *mean* are also often used by speakers of other varieties of English, but the use of the verb *feel* to express an opinion or a belief seems to be a specific feature of Indian English speakers. Consider the following sentences for illustration.

(4.97) a. But what I feel is <,,> as you said the funds <,,> are misused <,>
(ICE-IND:S1A-020)

b. What I feel is <,> only the education and the <,> economical status makes a woman to feel very liberated <,> (ICE-IND:S1A-011)

c. What I feel is <,> that food and food habit <,> do affect our mind <,> thought (ICE-IND:S1A-072)

According to the OED Online, the use of the verb *feel* to express a belief or an opinion goes back to the fourteenth century but is now obsolete (cf. "feel, v." OED Online; entry II 15 a; accessed 03/12/2014). The present data suggest that this usage of the

57 The *do*-type of pseudo-cleft seems to be common in speech and writing in Canadian English. Analyzing precisely this type of pseudo-cleft, Mair and Winkle (2012) find that the construction is most frequent in Canadian English among the ten varieties of English they investigate. Their study is based on both the spoken and written parts of ICE.

verb *feel* seems to have survived in the speech of Indian English speakers, though.

The high proportion of only a small number of specific verbs in Indian English suggests that in this variety of English - more so than in the other varieties - the cleft clauses seem to be used as formulaic expressions and to have a chunk-like character. This impression is underlined by the fact that there is often a pause after the copular (cf. the mark-up <,> and <,,> in the examples above) or the copular *be* is omitted and a pause occurs in its place, making the link between cleft clause and clefted constituent somewhat looser. Note, furthermore, that the expression *what I mean is* has been identified as a characteristic feature of Indian English in the literature and it has been claimed that it is used as a filler or for hedging (Sailaja 2009). In the present study, it is not so much the verb *mean* alone that is used in this sense - it is used more often by Singapore English speakers, for example - but rather a group of verbs including, in addition to *mean*, the verbs *think* and *feel*.

Summary

Summing up the investigation of basic pseudo-clefts, it can be noted that there is quite some variation in terms of frequency of use across the varieties of English analyzed, with Irish English and Hong Kong English showing the lowest frequencies. While cleft constructions are generally rare in the latter variety it is surprising that pseudo-clefts are so relatively infrequent in Irish English. The highest frequency of basic pseudo-clefts can be attested for Jamaican English.

Furthermore, there is some variation in terms of pseudo-cleft type. While WH-clefts are generally the preferred variant, Irish English speakers stand out in that they use TH-clefts and ALL-clefts to roughly the same extent as WH-clefts. In Indian English, on the other hand, WH-clefts account for nearly all of the basic pseudo-clefts. The examination of the verbs in the cleft clauses revealed that there is a small group of verbs that is frequently used by Indian English speakers, which suggests that the construction is used as a rather fixed expression in this variety of English. It has a more chunk-like character in Indian English and is often used as an utterance launcher, especially when speakers want to express their opinion. In Jamaican English, on the other hand, we find much more variation with respect to verb type in the cleft clauses. This suggests that WH-clefts are used more flexibly and for more purposes in this variety of English, which in turn may contribute to the overall higher incidence of basic pseudo-clefts in Jamaican English.

4.5.4 Reversed pseudo-clefts

Reversed pseudo-clefts are much more frequent than the other two types of cleft construction, as noted earlier. This has mainly to do with the type of construction exemplified in (4.98).

(4.98) a. That's what he says (ICE-PHI:S1A-003)

b. That's how I'm feeling (ICE-CAN:S1A-034)

c. This is why I've been mad at him (ICE-CAN:S1A-011)

This type of reversed pseudo-cleft is introduced by a demonstrative pronoun, usually *that*, rarely also *this* or *those*. These demonstrative clefts, as they are sometimes also called in the literature, will be the subject of the following subsection.

Demonstrative clefts

In previous studies, demonstrative clefts have been identified as the most common type of cleft construction in conversation. Biber et al. (1999: 961), for example, find that demonstrative clefts are sharply stratified by register. They are the by far most frequent type of cleft construction in their conversation data but occur rarely in academic prose. Similarly, Calude (2008: 78) finds that demonstrative clefts "constitute the most frequent cleft type" in her data of spontaneous spoken New Zealand English (47% of all cleft constructions). She argues that the construction is so well suited to spoken interaction because it is low in information content and cognitive load. That is, demonstrative clefts "require little or no planning since they typically involve a given/inferable cleft clause and a deictic cleft constituent, and can be used at the immediate discourse level to point to recently mentioned parts of discourse" (Calude 2008: 107). In line with these previous studies, demonstrative clefts also constitute the most frequent type of cleft construction in the present data. Table 4.24 gives the absolute token frequencies of demonstrative clefts (demCs), their proportions out of all clefts and out of all reversed pseudo-clefts (revPCs).

Table 4.24: Demonstrative clefts (absolute token frequencies, percentages out of all clefts and percentages out of all reversed pseudo-clefts).

	GB	IRE	NZ	CAN	SIN	PHI	JAM	IND	HK
demCs	198	241	265	254	194	236	236	194	107
% out of clefts	46.3	53.6	56.3	56.7	55.3	48.2	44.4	51.3	55.2
% out of revPCs	81.1	89.3	88.3	85.5	87.0	76.9	79.7	93.7	92.2

As can be seen, demonstrative clefts make up more than half of all cleft constructions in most of the data. The percentages are below 50% only in the British English, Philippine English and Jamaican English data. This is due to a combination of two factors. First, basic pseudo-clefts are relatively more frequent in these varieties of English and, second, 'proper' reversed pseudo-clefts - that is, reversed pseudo-clefts without initial demonstrative pronouns - also make up slightly higher proportions, as the percentages in the last row (% out of revPCs) show.[58]

The sentences in (4.98) illustrate that different *wh*-words may occur in the cleft

58 The distribution of the different cleft types (i.e. IT-clefts, basic pseudo-clefts, 'proper' reversed pseudo-clefts and demonstrative clefts) within each variety of English analyzed is not random. Chi-

clause. According to Biber et al. (1999: 961), demonstrative clefts with interrogative pronouns *what* and *why* are particularly common, occurring roughly 50 and 20 times per 100,000 words, respectively, in their data. The pronouns *how*, *where* and *when*, on the other hand, are considerably less common, with only 5 tokens per 100,000 words. As for the present study, the frequencies of these five *wh*-words in demonstrative clefts are presented in Table 4.25 (frequencies per 100,000 words).

Table 4.25: Distribution of *wh*-words in demonstrative clefts (frequencies per 100,000 words).

corpus	what	why	how	where	when
GB	44.6	19.3	6.9	7.9	3.5
IRE	56.5	14.4	5.9	11.9	5.5
NZ	55.7	20.9	3.9	12.6	2.6
CAN	59.8	18.5	4.3	11.9	5.7
SIN	29.0	50.2	3.9	3.4	3.4
PHI	38.3	44.3	6.5	6.0	3.2
JAM	54.9	16.9	11.3	9.8	3.8
IND	35.2	41.2	6.9	0.9	0.5
HK	12.2	27.3	1.3	0.4	-

As can be seen, there is quite some variation across the varieties of English analyzed. In line with Biber et al. (1999), demonstrative clefts with *what* and *why* are the most common variants. In the L1 English varieties and Jamaican English, they occur roughly 50 and 20 times per 100,000 words, respectively, like in Biber et al.'s data. However, in the L2 English varieties Singapore English, Philippine English, Indian English and Hong Kong English it is the *why*-type that is more common than the *what*-type. Furthermore, it can be observed that demonstrative clefts with *where* are much more frequent in some varieties in the present study than the 5 tokens per 100,000 words that Biber et al. report. In Irish English, New Zealand English and Canadian English they occur around 12 times per 100,000 words. In Indian English, on the other hand, demonstrative clefts are largely restricted to the patterns with *what*, *why* and *how*; in Kong Kong English it is even only the two former types that occur in substantial numbers.

In addition to these five *wh*-words, it is interesting to note that some instances with *who* have been found in the data: five tokens in Irish English, two in New Zealand English and one token in British English. Examples are given in (4.99).

(4.99) a. Well that's who I played with over Christmas <,> in the Maltings <,> in Aldeborough (ICE-GB:S1A-058)

b. So that's who i talked to essentially (ICE-NZ:S1A-099)

c. That's who Mam's just talking about (ICE-IRE:S1A-078)

squared tests, comparing the frequencies of the different cleft types in each corpus, yield highly significant results in all nine cases ($p < 0.001$).

While the pronoun *what* is also almost exclusively found in basic pseudo-clefts, the pronoun *who* is not permitted at all. That is, the basic pseudo-cleft counterpart to sentence (4.99b) is not admissible: **Who I talked to essentially is that.* As pointed out in the section on basic pseudo-clefts, in these cases a pro-noun is used instead, for example, *the one* or *the person*: *The person I talked to essentially is that.* This might be the reason why it is surprising to find demonstrative clefts with *who* in the data. They seem to be a very rare phenomenon, however. They are not mentioned by Calude (2008) or Biber et al. (1999) and there are only eight items in the present study.

After this brief discussion of demonstrative clefts, the following subsection will examine reversed WH-clefts, ALL-clefts and TH-clefts.[59]

Types of reversed pseudo-cleft

Reversed pseudo-clefts can also be categorized into three different types, depending on the initial element of the cleft clause. As in the case of basic pseudo-clefts, WH-clefts, ALL-clefts and TH-clefts can be distinguished. Examples of these three variants are given in (4.100).

(4.100) a. Friedrichstor was **where** the hotel was (ICE-JAM:S1A-072)

b. She was **all** I looked at (ICE-PHI:S1A-019)

c. Mia is **the one** I remember (ICE-IRE:S1A-070)

These sentences constitute instances of 'proper' reversed pseudo-clefts. I use the term 'proper' here simply for convenience in order to have a label which allows me to refer to those tokens of reversed pseudo-cleft that do not have an initial demonstrative pronoun. The examples of demonstrative clefts given so far have all been instances of reversed *wh*-clefts, but there are, of course, also items with *all* and pro-nouns:

(4.101) a. That's **all** you need (ICE-IRE:S1A-048)

b. That's **the thing** you make the most money out of (ICE-IRE:S1A-010)

The distribution of these three types of reversed pseudo-cleft is represented in Figure 4.24.[60] As can be seen, the variation across the varieties of English analyzed is not as pronounced as in the case of basic pseudo-clefts. Reversed ALL-clefts occur very rarely in all nine varieties of English analyzed. As noted earlier, they are most frequent in Irish English with 18 tokens (6.7% of all reversed pseudo-clefts), followed by New Zealand English with 16 tokens (5.3%). The great majority of reversed pseudo-clefts have a *wh*-word as the initial element in the cleft clause. The proportions range from 69.7% of all reversed pseudo-clefts in Philippine English to 88.4% in Indian English. That is, like in the case of basic pseudo-clefts, the preponderance of the *wh*-type

59 The present study examined only a small number of syntactic features of demonstrative clefts. For a comprehensive discussion of the construction's discourse functions see Calude (2008).

60 The token frequencies of the different types of basic pseudo-cleft and their proportions within the ICE samples are given in Appendix 6.10.3.

is again most dominant in Indian English. Interestingly, way over 90% of reversed WH-clefts are demonstrative clefts in all nine varieties; in Singapore English and Indian English it is even all WH-clefts that are demonstrative clefts.

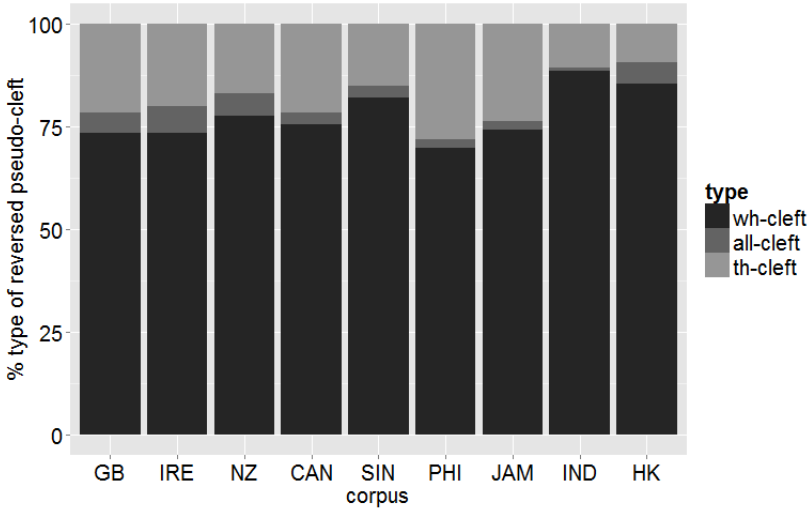

Figure 4.24: The proportions of reversed WH-clefts, ALL-clefts and TH-clefts (percentages out of all reversed pseudo-clefts).

Reversed ALL-clefts are also almost exclusively demonstrative clefts in all varieties of English. With TH-clefts the picture is somewhat more varied, however. In the L2 English varieties and British English, TH-clefts are most frequently of the 'proper' type, as exemplified in sentence (4.100c); they are even categorical of the 'proper' type in the case of Singapore English. In the other L1 English varieties, on the other hand, TH-clefts are slightly more often demonstrative clefts, such as the sentence in (4.101b).

Philippine English speakers seem to prefer one specific type of reversed TH-cleft in particular. In their speech - more so than in the speech of the other speakers - there are many instances of 'proper' reversed pseudo-clefts with the pro-noun *the one*:

(4.102) a. You were the one who had to talk (ICE-PHI:S1A-057)

b. Thomas was the one who was assigned here *ha* (ICE-PHI:S1A-028)

c. The students are the ones who make their their careers (ICE-PHI:S1A-082)

In the Philippine English data, there are 57 sentences of this type, which account for 18.6% of all reversed pseudo-clefts. This also explains the slightly higher proportion of reversed TH-clefts in this variety of English as compared to the other varieties.

Coming back to Indian English, it has been noted above that the *wh*-type is most dominant in this variety of English and that all of the tokens are demonstrative clefts. This finding ties in with the results of the analysis of basic pseudo-clefts, where it was observed that WH-clefts are the predominant variant and that this may be due to the fact that some constructions have become fixed expressions. The same seems to hold for reversed WH-clefts. Recall that in Indian English almost all demonstrative clefts

have *what* or *why* in the cleft clause, occasionally also *how*. This suggests that again a few fixed types are preferably used.

Summary

The analysis of reversed pseudo-clefts has revealed some variation in terms of frequency. While the construction is the most common type of cleft in all nine varieties analyzed, its frequencies are much smaller in Singapore English, Indian English and Hong Kong English than in the other varieties of English analyzed. It has been observed that demonstrative clefts make up the great majority of reversed pseudo-clefts in all varieties, with their preponderance being most dominant in Indian English and Hong Kong English. As far as the *wh*-words in the cleft clause of demonstrative pronouns is concerned, it can be noted that *what* and *why* are the most frequent variants in all varieties, but it is interesting to observe that L1 English speakers and Jamaican English speakers predominantly use *what* while it is *why* that is more frequent in the speech of the other L2 English speakers. The preponderance of *what* and *why* over *how*, *where* and *when* is particularly striking in Indian English and Hong Kong English.

In Indian English, demonstrative clefts of the *wh*-type with pronouns *what* and *why* are the preferred variant, which underlines the impression that pseudo-clefts (both basic and reversed) are used as rather formulaic expressions. In the other varieties of English we find a similarly strong preference of demonstrative WH-clefts, but the range of different structures seems to be particularly restricted in Indian English.

The proportions of WH-clefts, ALL-clefts and TH-clefts are very similar across the varieties of English analyzed. That is, ALL-clefts are rare in all varieties and WH-clefts make up the large majority. For Philippine English, a somewhat stronger preference of reversed pseudo-clefts containing the pro-noun *the one* can be attested.

4.5.5 Concluding remarks

The discussion of the three major types of cleft construction in this section has concentrated on structural similarities and differences among the varieties of English analyzed, but what is certainly also worth examining is the discourse functions of the constructions. Differences in frequency are very likely also motivated by the different pragmatic functions the constructions may serve. The distribution of old and new information varies across the three types of cleft, but what all of them have in common is that they are specificational sentences that put emphasis on the clefted element, although to different degrees. IT-clefts, basic pseudo-clefts and reversed pseudo-clefts also differ in the way they are used in discourse. These pragmatic differences appear to be determined by the information-structural tendency to place topics near the beginning of a sentence. That is, an element will preferably be put at the beginning of a clause if it continues the topic of the preceding discourse (e.g. Halliday 1967; De-

clerck 1984). Consider the examples in (4.103), with a pseudo-cleft in (a), an IT-cleft in (b) and a reversed pseudo-cleft in (c).

(4.103) a. Have you found everything you need? - Well, I've found the handbooks that I need, but **what I haven't found is the dictionary**.

b. Have you found everything you need? - Well, I've found the handbooks but **it's the dictionary that I haven't found**.

c. Have you found everything you need? - Well, I've found the handbooks but **the dictionary is what I haven't found**.

(Declerck 1984: 275)

The reply in (a) is felicitous because the cleft clause of the pseudo-cleft *What I haven't found* picks up the theme of the question (whether I have found everything I need). Thus, this piece of conversation sounds much more natural than that in (b) or (c). In the latter two sentences, the clefts begin with the clefted constituents *it's the dictionary* and *the dictionary*, which suggests that the conversation is about the dictionary and not about what I have or have not found. Thus, the information flow is somehow broken because the reply does not really fit the question. What the examples also illustrate is that in information-structural terms reversed pseudo-clefts are closer in their behaviour to IT-clefts than to basic pseudo-clefts. IT-clefts and ('proper') reversed pseudo-clefts typically begin with the new information, while basic pseudo-clefts begin with the old or known information. This may also explain why we find comparatively small frequencies of IT-clefts and reversed pseudo-clefts in Indian English but not of basic pseudo-clefts. Lange (2012) argues that a major discourse motivation for Indian English speakers is to create topic continuity, that is, they often pick up old information from the preceding discourse. The investigation of left dislocation and fronting constructions (cf. sections 4.1 and 4.3) shows that this discourse motivation is not exclusive to Indian English speakers but seems to be even more important for Philippine English speakers in the case of these two constructions. Yet, with respect to cleft constructions the creation of topic continuity is possibly a more important function for Indian English speakers. They frequently use basic pseudo-clefts and among the reversed pseudo-clefts almost exclusively demonstrative clefts, exactly those types of cleft that begin with old or known information. IT-clefts and 'proper' reversed pseudo-clefts, on the other hand, typically begin with the new information and are used far less frequently by Indian English speakers both in comparison to the other cleft types and in comparison to the speakers of the other varieties (with the exception of Hong Kong English speakers, who generally use cleft constructions infrequently).

Another aspect that may influence speakers' preferences concerns the length of the constituents. It seems that the shorter constituent tends to come first in the sentence. Consider the following sentences where (a) and (b) definitely sound more natural than (c).

(4.104) a. Who loves apples? - It's Tom who loves apples.

b. Who loves apples? - Tom is the one who loves apples.

c. Who loves apples? - The one who loves apples is Tom.

It seems, however, that the structuring of information is the stronger influencing factor on speakers' choices of the cleft type. Consider, for example, the TH-cleft in (4.105) taken from ICE-Jamaica.

(4.105) The only person who might have his own secretary is the director you know
(ICE-JAM:S1A-027)

In this sentence, we find a very long initial element *The only person who might have his own secretary* followed by a short highlighted element *the director*. If the length of the constituents was the stronger determining factor we would expect a cleft construction like *The director is the only one who has his own secretary* or *It's the director who has his own secretary*. The speaker's preference for this type of cleft can be accounted for by looking at the preceding discourse: the immediately preceding sentence is *So like one secretary would have to work with three <,> technical staff*. That means that the speaker places the topic of the preceding discourse at the very beginning of the utterance and thus organizes the information in a coherent way keeping up the information flow.

Given the variation in frequency and therefore in preferred cleft type across the nine varieties of English analyzed, it might be worth comparing these pragmatic functions and the use of the three major types of cleft construction across varieties of English. This may also provide a clearer picture of their interaction with other information-packaging constructions, as suggested by the findings for Indian English. I leave this for future research.

CHAPTER 5

Concluding discussion

The major aim of the present study was to examine structural information-packaging strategies in World Englishes and to provide a comprehensive view on their distribution and use. Canonical English sentences typically have the word order SVX, but sometimes speakers may deviate from this rather fixed word order for pragmatic purposes and shift elements in the clause into non-canonical positions. An important factor in this respect is the information status of the clause constituents, that is, the structuring of clauses very much depends on whether a constituent refers to discourse/speaker-old information or to discourse/speaker-new information. Thus, speakers may, for example, deviate from the canonical word order in order to introduce new information into the discourse or re-introduce an entity after a longer gap of absence. Another motivating factor involves the highlighting of certain clause constituents, that is, speakers may use specific structures to contrast one piece of information with another or to focus the addressee's attention on a certain piece of information. It was expected to find variation in terms of frequency of use of the different structural information-packaging constructions, in terms of the preferred realizational variants and speakers' motivations for their preferred usage patterns. It was furthermore expected that a high frequency of use might lead to the strengthening of a feature in the speaker's memory and possibly also to a widening of scope of realization options, which in turn might again lead to an increased frequency of use.

The major factors that were assumed to impact on the frequency of use of the different information-packaging constructions involved substrate influence in language contact situations, universals of language acquisition and specific features of the socio-cultural setting. Hence, in addition to providing a comprehensive view on the constructions at issue, the present study also aimed at offering new insights into the mechanisms of language contact and its outcomes. Furthermore, it was interested in finding out more about the interplay of language contact, universal learner and processing strategies and sociolinguistic and pragmatic factors. It was expected that these factors do not act in isolation, but rather that there was a network of interacting forces at play in the shaping of linguistic knowledge.

The investigation of the different information-packaging constructions - left and right dislocation, fronting, *there*-existentials, IT-clefts and pseudo-clefts - shows that speakers of different varieties of English indeed prefer different strategies to structure the information in a sentence, motivated by different - at times interacting - forces. Take, for example, left dislocation constructions. We have observed that they are gen-

erally more frequent in the speech of L2 English speakers than in that of L1 English speakers, a finding which supports previous studies claiming that left dislocation is a characteristic feature of the speech of learners of English (cf. e.g. Gruber 1967; Chambers 1973; Cotton 1978; Williams 1987; Mesthrie 1992; Carter/McCarthy 1995; Ortega 2009). The overwhelmingly higher incidence of the construction in Indian English, however, suggests that there must be other motivating factors as well. Since the investigation of a number of syntactic and pragmatic features yields no conclusive results that can plausibly account for the high frequency, it is concluded that left dislocation constructions are pragmatically less marked for Indian English speakers than for speakers of the other varieties. The very high proportion of subject left dislocation tokens suggests that the construction is predominantly used as a topic-establishment device, with the co-referential pronoun serving as a topic marker. It is very likely that influence from the Indian background languages plays an important role in this respect. As noted earlier, an important underlying assumption of the present study is that multilingual speakers have a pool of linguistic features at their disposal from which they can choose freely. It is furthermore assumed that features of the different languages in contact influence one another. In Malayalam, for example, it is sufficient to mark a constituent as the topic by placing it in sentence-initial position, just like in English. Yet, it is also possible to use a more explicit marker, which is placed after the topic. That is, in such cases, "the topic remains in first place in the sentence but is followed by a reinforcing element" which is attached to the topic (Asher/Kumari 1997: 184). Similarly, in Hindi the particle *to* may be used to mark the topic of a sentence. Given these structures in the Indian background languages, Indian English speakers may simply mark the subject as the topic by using the co-referential pronoun as a "reinforcing element", a strategy they are familiar with from their L1. I would thus suggest that different forces interact to unmark the LD construction and increase its usage in Indian English. For one, the feature pool of Indian English speakers contains the possibility of marking topics explicitly by means of an ending that is attached to the topic or by means of a particle. Additionally, they are familiar with the left dislocation construction as a means of establishing a topic. Since morphological marking is not admissible in English, Indian English speakers do not directly transfer the feature from their L1 into English but a construction that is already there assumes this function, that is, the co-referential pronoun is used as an explicit topic marker. The construction may therefore be used more frequently and it may be less marked than in, for example, British English. It seems, then, that in Indian English left dislocation constructions tend to be used in a similar way as, for example, in French, where the construction occurs very frequently but without any specific pragmatic load.

Coming back to the issue of speakers' preferred structures and their motivations, we have seen that in the case of left dislocation constructions there are different forces at work: their simplifying function (the breaking down of an utterance into smaller chunks makes production and processing easier) makes them a preferred structure among learners of English and, additionally, influence from the background languages has very likely led to an increasing use and the unmarking of the construction in Indian English.

A similar process of unmarking of a marked structure can be witnessed for fronting constructions. The results of the present study illustrate that Indian English speakers

also use this structure overwhelmingly more frequently than speakers of the other varieties analyzed. It has been observed that it is typically objects that are placed in initial position, which suggests that for Indian English speakers the word order OSV is possibly less marked than for speakers of other English varieties. It is very likely that it is again the Indian background languages that play an important role in this respect. In the case of fronting constructions this means that the basic SOV sentence structure of Indian languages and their much more flexible word order impacts on the structuring of sentences in Indian English. Consequently, the fronting of objects is less marked in Indian English and the word order OSV is maybe even on its way to becoming one of the basic sentence patterns for Indian English speakers (see the example from Malayalam below).

The unmarking of a pragmatically marked structure is in fact fairly common, as Heine (2008) notes in an article on word order:

> But perhaps the main driving force for adjusting one's word order to that of another language is to select a pragmatically marked use pattern that exhibits an ordering corresponding to that of the model language and to grammaticalize that pattern into an unmarked syntactic pattern; note that a development from pragmatically marked to syntactic constituent is a fairly common grammaticalization process (Heine 2008: 43)

Heine reports on a case of word order change in Arabic where "a topicalization strategy within VO syntax was grammaticalized to a pragmatically unmarked OV syntax" (2008: 51). The 'topicalization strategy' in this case involves left dislocation which allows the OV order of the model languages Tajik and Uzbek while also adhering to the VO order (the resumptive object pronoun comes after the verb). What is different in the case of Indian English is that objects are frequently fronted without placing a co-referential pronoun in the core of the clause. But the following example from Malayalam illustrates that this is in fact what is done in this Indian language.

(5.1) *puuccaye* *ellaarum* *kuuti* *talli* *konnu*
cat-ACC all together beat-PP kill-PAST
'The cat, they all beat it to death.'

(Asher/Kumari 1997: 184)

Asher and Kumari (1997) note that due to the free movement of constituents in Malayalam other elements than the subject are allowed to occupy the topic slot. Interestingly, the authors insert a co-referential pronoun in the English translation although it is obviously not present in the Malayalam sentence, as the literal gloss indicates. I find this somewhat puzzling because given the results of the present study, Indian English speakers would probably rather say *The cat they all beat to death*, just like in the Malayalam sentence. Maybe this example indicates that left dislocation and fronting constructions are very similar phenomena in Indian languages.

Another factor that may contribute to the high incidence of fronting constructions in Indian English is that the construction might serve functions for which speakers of other varieties would rather use the IT-cleft construction. The results of the present study show that IT-clefts are rarely used by Indian English speakers in comparison to speakers of the other varieties, which suggests that they prefer fronting constructions for emphatic and contrastive purposes, the functions that are typically associated with

the IT-cleft construction. The Indian background languages may again play an important role because to my knowledge there is no construction comparable to the English cleft. In Hindi, an element under focus can be identified by emphatic stress, focus particles or movement from its canonical position. Similarly, in Malayalam deviations from the canonical word order SOV allow the highlighting of a constituent. Indian English speakers may therefore unconsciously prefer fronting constructions over cleft constructions because the latter do not exist in their L1.

The investigation of left dislocation and fronting constructions has thus shown that high frequencies may lead to the strengthening of a construction's representation in the speakers' memories and to its pragmatic unmarking. What is interesting to note about Indian English speakers is that they predominantly use 'default' tokens of both constructions, that is, left-dislocated subjects and fronted objects.[1] It seems that subject left dislocation and object fronting are constructions that are so deeply entrenched in the Indian English speakers' minds that their activation has become an automated routine.

Another possible scenario that may follow from the high frequency and deep entrenchment of a construction is that its scope of realizational variants or its range of functions is widened. Take, for example, left dislocation or IT-cleft constructions in Irish English. In the case of left dislocation, it is interesting to observe that it is used more frequently by Irish English speakers than by the other L1 English speakers. For one, this may be due to the fact that from a historical point of view Irish English can be described as a L2 English variety because "it has evolved as a result of long-standing coexistence and contacts with the indigenous Celtic language of the Irish people, Irish" (Filppula 2012: 31). As noted earlier, left dislocation is generally found more frequently in the speech of L2 English speakers. Additionally, influence from Irish has probably led to a higher incidence of the construction and a wider scope of realizational variants. The type of situation in which the native Irish acquired English is important to consider in this respect because there was only little if any formal education for the majority of the Irish population. This led to an unguided or even uncontrolled way of language acquisition, with hardly any restrictions on non-standard features stemming from Irish influence (cf. Hickey 2007: 125). This means that today there are many features in Irish English which are very likely influenced by Irish. A case in point is the use of possessive left dislocation tokens, whose somewhat higher incidence in Irish English might well be due to Irish because in this language we find resumptive possessive pronouns in the passive and the progressive constructions. Hence, the possessive resumptive pronoun may be more readily available to Irish English speakers. Similarly, in Irish English IT-clefts, the clefted elements serve a greater variety of syntactic functions than in the other varieties of English analyzed, which might also be due to Irish where clefts are used much more flexibly than in English. Furthermore, subordinator choice in IT-clefts seems to be influenced by Irish, with subordinators in subject position often being omitted. Similarly, in *there*-existentials with relative clause extensions the somewhat higher incidence of subject-zero relatives in Irish English may be due to influence from Irish.

1 Left-dislocated subjects and fronted objects are also the most common types among the speakers of the other varieties of English analyzed. That is why they are called the 'default' cases. In Indian English, however, their preponderance is particularly striking.

In sum, the results of the present analysis indicate that we have two very different situations: while high token frequencies seem to have resulted in the entrenchment of a few types of construction in the minds of Indian English speakers, the high token frequencies in Irish English (especially in comparison to the other L1 English varieties) are accompanied by a wider range of realizational variants. It is difficult to say whether high frequency is the cause or the effect of a deeper entrenchment and a wider range of realizations in the latter case. Both scenarios are possible: high token frequencies may lead to a broadening of the range of a construction's variants and its functions, but it may also be the other way round and high token frequencies may result from a larger range of variants and functions. A third possible scenario can be described as 'circular', with frequency and strengthened representation interacting and influencing one another. This finding reminds us of the fact that frequency can be cause or effect of certain developments in linguistic structure, or both at the same time.

In addition to Indian English with its multilingual speakers and Irish English with its strong Celtic influence, valuable data for the present study also come from those English varieties whose speakers have a topic-prominent L1 because basic sentences in topic-prominent languages are structurally very different from those in the (subject-prominent) English language and are very likely to impact on the latter. Particularly interesting is the comparison of the speech of Singapore English and Hong Kong English speakers because most of them have a Chinese dialect as their L1, but they clearly differ in terms of their proficiency in English. Furthermore, Singapore and Hong Kong differ in terms of their socio-cultural setting or the role they ascribe to the English language. In Singapore, English is the "language for the construction and expression of the Singaporean (i.e. national) identity" (Lick/Alsagoff 1998: 207). The importance the government ascribes to the language is seen, for example, in the education system, with English being the medium of instruction. Furthermore, English is becoming a home language for more and more speakers. In Hong Kong, on the other hand, English is rarely used in daily life. In fact, the question of whether there is an (semi-)autonomous variety of Hong Kong English is still being debated among experts. Hence, the status of English in these two territories is very different and this, in turn, certainly also impacts on the use of and proficiency in the language. Taking all these factors into account, the comparison of Singapore English and Hong Kong English data may be particularly useful for differentiating between learner features and contact-induced developments.

A first interesting finding in this respect is the low frequency of cleft constructions in the Singapore English and especially the Hong Kong English data. Influence from the Chinese background languages may, at least in part, be responsible for that. For one, relative clauses are structured differently in Chinese than in English (pre- vs. postmodifying structures), which might make Chinese learners of English shy away from using complex English structures which involve relative clauses (cf. Chan 2004). Second, Chinese is a pronoun-dropping language and does not syntactically require dummy subjects. Hence, the *it* in IT-clefts has no equivalent in Chinese (cf. Li/Thompson 1981: 91; Huang 1984). Third, in English IT-clefts, the focused element is moved to the periphery of the clause and placed after *it be*. The Cantonese *haih…ge* and the Mandarin *shì…de* structures, on the other hand, involve no such word or-

der change. Taking these three factors together, this might be the reason why only a relatively small number of clefts can be attested for Singapore English and Hong Kong English. Because of the different proficiency levels of the speakers in these two countries the influence from the Chinese language is particularly strong on Hong Kong English.

Another finding that illustrates the influence of the topic-prominent background languages concerns fronting constructions. It is interesting to note that fronted prepositional phrases are (nearly) absent from Singapore English, Hong Kong English and Philippine English, exactly those three varieties of English that have topic-prominent background languages. Since prepositions are often omitted in these (and other) L2 varieties of English, (possible) fronted prepositional phrases can easily be confused with 'hanging topic' constructions. They have therefore been excluded from the present analysis. The occurrence of 'hanging topic' constructions has been described in the literature as typical of topic-prominent languages (Lambrecht 2001a; Setter et al. 2010), but they obviously do also occur in subject-prominent languages such as English. Hence, it might be worth analyzing 'hanging topic' constructions in varieties of English in order to see whether topic-prominent background languages have an effect in terms of frequency of use and whether or in what way they interact with fronting and left dislocation constructions.

Exactly these three varieties - Singapore English, Hong Kong English and Philippine English - stand out among the other varieties of English analyzed in another respect, namely in that the preposition *for* has grammaticalized into a topic marker (e.g. *For me I like badminton*). This *for*-LD construction, as it is called in the present study, is often used to express an opinion or attitude and often in contrast to some other person's opinion. *For* was also used for the establishment of a topic in earlier English, modelled on French *pour moi*, but it was later abandoned, according to the Online Oxford English Dictionary. It is interesting to note that *for* has again grammaticalized into a topic marker in some varieties of English. The topic prominence of these speakers' L1s very likely plays an important role in this process of (contact-induced) grammaticalization. These English speakers are more sensitive to the notion of topic and attempt to structure sentences according to the principles of their L1. In addition, these speakers' pool of linguistic features contains the *as for*-construction, which is commonly used in English to establish a topic (e.g. *As for books, I like Edgar Allan Poe*), and the *for NP*-construction, which is commonly used to express an opinion or attitude, often in contrast to some other person's opinion or attitude (e.g. *For me it is quite a bore*). In sum, I would suggest that these three features - topic prominence of the L1 and the *as for*- and the *for NP*-constructions of English - have contributed predominantly to the grammaticalization of *for*.

The *for*-LD construction is most frequent and shows the greatest variation in the Hong Kong English data. This is probably due to the somewhat lower proficiency level of Hong Kong English speakers, especially in comparison to Singapore English speakers, and their higher susceptibility to structuring sentences according to the principles of their L1. In GloWbE, *for*-LD constructions occur only very infrequently, but the same trends can be observed, especially with respect to the type where *for* is followed by a noun rather than a pronoun. The low frequency of the construction in GloWbE furthermore suggests that it is really a feature of spontaneous spoken inter-

action.

In sum, then, the findings of the present study suggest that *for* is used as a topic marker predominantly in those varieties of English which are in contact with topic-prominent languages. Since the frequencies are overall rather small, these claims clearly call for further research based on larger datasets. Moreover, it would be very interesting to conduct follow-up studies on more recent data to see whether the *for*-LD construction has spread further and to see whether we are here dealing with a case of ongoing grammaticalization. It will furthermore be interesting to see whether the construction is really largely restricted to those varieties of English whose speakers have a topic-prominent L1 or whether they are only in the lead of a general development, spreading through all varieties of English. Such a general development would be similar to the change that has been observed for the construction 'as far as X is concerned/goes' where the coda is today often omitted (cf. Rickford et al. 1995; Britain 2000).

Influence from the topic-prominent L1 has also been noted in the literature to play a role in the formation of relative clauses. More precisely speaking, Chinese learners of English have been reported to frequently omit relative markers in subject position. In particular, the omission of subject relatives in *there*-existentials with relative clause extensions has been described as a characteristic feature of Chinese learners of English (Schachter/Celce-Murcia 1977; Newbrook 1988, 1998; Li 2000; Ortega 2009; Hung 2012). The results of the present study confirm these claims, with Hong Kong English showing the highest proportion of subject-zero relatives in *there*-existentials among the varieties of English analyzed. The data suggest that this is again a learner feature influenced by the topic-prominent L1: Hong Kong English speakers attempt to structure sentences according to the principles of their L1, giving the topic first followed by the comment. This impression is underlined by the fact that in Hong Kong English existentials with subject-zero relatives typically have a very loose syntax and contain further learner features (e.g. lack of plural marking or the placement of adverbs: *In my school <,> there's so many student like uh bad behave*). Since subject-zero relatives in existentials are much less frequent in Singapore English, we are here again dealing with a situation that can be accounted for by the somewhat lower proficiency level of Hong Kong English speakers.

Singapore English speakers' higher proficiency and the institutionalization of this variety of English can additionally be seen in that it has already developed some local norms. An example from the present study is the use of *got* to express existence (e.g. *Cake inside got fruits ~There is fruit in the cake*). The construction has very likely inherited its uses from Chinese *yǒu* (cf. Bao 2014). Speakers' creative innovations may only result in language change if they are accepted in the speech community, as Matras and Sakal (2007) point out:

> [...] learners' innovations may result in long-term change, but only in situations in which the learners constitute a large enough collective and the process of language acquisition never actually 'catches up' with the model or native form of the target language - the classic 'substrate influence' scenario. Replication of an external model will only lead to change if normative control within the speech community is relatively lax and flexible enough to allow a drift toward regularisation and acceptability of the new imported structures. (Matras/Sakel 2007: 849)

The innovative use of *got* to express existence provides evidence that normative control in Singapore English is relatively lax and flexible and allows for local norms to develop.

With regard to the omission of relative markers/subordinators in *there*-existentials and IT-clefts, the Irish English data yield interesting results as well.[2] This variety of English shows the second highest proportions of subject-zero relatives/subordinators in both cases, following Hong Kong English. While the topic-prominent L1 and the low proficiency of its speakers are very likely responsible for the high incidence of these constructions in Hong Kong English, in Irish English the omission of subject-zero relatives/subordinators can well be traced back to Irish. A striking difference was noted in the realization of the constructions: while we find a loose syntax and other learner features in the Hong Kong English sentences, the structures are much more compact in Irish English, which suggests that the structures are well entrenched in the latter speakers' minds.

In sum, the analysis of subordinator or relative marker choice in *there*-existentials and IT-clefts again shows that various factors impact on the shaping of linguistic knowledge and that different influencing factors may lead to similar outcomes: the topic prominence of Hong Kong English speakers' L1 and the sentence structure of Irish both lead to a higher incidence of subject-zero relatives/subordinators in the relevant varieties of English. Furthermore, the analysis shows that it is not easy to disentangle the network of forces that impact on the emergence of linguistic patterns. In the present case, we have language contact or transfer and acquisitional factors that play an important role, but it seems that it is both factors together (and possibly others) that are responsible for the observed usage patterns.

Methodologically, the present study shows that the complementing of ICE data with larger datasets is a useful set-up. The ICE family provides valuable data as the individual components have a common design and thus provide comparable data of a wide range of different varieties of English. Furthermore, the ICE corpora contain transcriptions of informal conversations, an invaluable set of data for the present study, whose research topic involves a number of marked structures which are predominantly found in spoken interaction. Yet, the ICE corpora have their limitations, as was repeatedly noted throughout the present study. The greatest disadvantage is probably their limited size, which is particularly relevant for the present study as the 'direct conversation' files of ICE the study is based on contain only about 200,000 words each. On the one hand, this was a feasible amount of text to read through for the manual annotation of the constructions at issue in the present study. But on the other hand, due to the limited size of the datasets token frequencies are often too small to allow for reliable conclusions to be drawn or to say whether the findings are really representative of the speech community. Furthermore, the results of statistical tests are obviously not as reliable as they would be if the datasets and token frequencies were larger. To compensate for the limitation of size, a number of ICE searches have been complemented by searches in GloWbE, the Corpus of Global Web-based English, which contains extremely more data per variety of English than ICE: the British English component of GloWbE, for example, contains more than 387 million

2 Recall that the term 'subordinator' is preferred over 'relative marker' in the case of IT-clefts in the present study because the status of the cleft clause is still debated in the literature (cf. also section 3.2.5).

words, the Irish English data comprise more than 101 million words and the Singapore English data nearly 43 million words. In the present study, complementing the ICE data by GloWbE proves to be particularly useful in the case of *for*-LD constructions. The GloWbE data provide further evidence in support of the claim that the construction is predominantly found in Singapore English and Hong Kong English. What the additional use of GloWbE also shows is that the *for*-LD construction seems largely to be a feature of spontaneous spoken language rather than the language of blogs and discussion forums because it occurs very rarely in GloWbE, as noted earlier. From this it must be concluded that, while the combination of ICE and GloWbE is a very useful research set-up, it also has its limitations when it comes to features that are characteristic of informal face-to-face interactions or telephone conversations. The data included in GloWbE, although to be seen at the interface between speech and writing, are still too different from informal spoken interaction to allow for more items to occur.

Another methodological issue that deserves mentioning concerns the formality levels and thus the compatibility of the ICE spoken texts. Despite the ICE guidelines there are differences in the texts or rather recordings that have been included in the 'direct conversation' files of ICE, resulting in different levels of formality. For the recording of informal speech of L2 English speakers, the main difficulty is, of course, that the majority of these speakers would normally not use the English language in informal settings but rather their mother tongue or home language. This may automatically lead to a somewhat higher level of formality in the spoken L2 English data, also because English is usually learned in a more formal setting at school and is associated with formality by these speakers. The discrepancy in formality is, however, not restricted to L2 English varieties but can also be found among the L1 English varieties. The 'direct conversation' files of ICE-Great Britain, for example, contain many interviews or conversations between doctor and patient or professor and student, which constitute much more formal settings than the many conversations among friends or family members that are part of ICE-Ireland and ICE-New Zealand. In the present study, this discrepancy in formality may well be responsible for one or the other deviant behaviour attested for British English as opposed to the other L1 English varieties. Take, for example, variable concord in existential *there*-constructions, with British English speakers using items with singular concord far less frequently than the other L1 English speakers (e.g. *There's little benches outside*). Recall that singular concord is usually associated with more informal conversations. The analysis of pronominal left dislocation (e.g. *Us we make good music*) seems to give further evidence of the more formal character of the 'direct conversation' files in ICE-Great Britain because the construction is surprisingly infrequent in the British English data, although previous studies claim otherwise (cf. e.g. Lambrecht 1994).

Some variation in the make-up of the 'direct conversation' files also seems to be present in the case of ICE-Canada. The present study finds various deviant usage patterns among Canadian English speakers which cannot reasonably be explained in syntactic or pragmatic terms. One example is the right dislocation construction, which occurs very infrequently in the Canadian English data, especially in comparison to the other L1 English varieties. Another example can be found among fronting constructions, with fronted complements being practically absent from the Canadian

English data, while they occur frequently in the other L1 English varieties. A common denominator of these two structures - right dislocation and fronted complements - is that they are often used to express feelings, an opinion, attitude, evaluation or judgement. Possibly, the 'direct conversation' files in ICE-Canada simply give not as many opportunities to talk about feelings and attitudes as is the case in the other L1 English data.

The present study set out to examine structural information-packaging strategies across World Englishes and to provide a comprehensive view on their distribution and use. The hypothesis that speakers have different preferences in the way they structure the information in a sentence or in the strategies they choose for highlighting or contrasting information in a sentence can be confirmed by the findings of the present study. We have seen that these differences are not categorical in nature, but we rather find a gradient from more frequent to less frequent usage of a construction. As for the motivating factors for the preferred structures, the present study aimed at shedding more light on the interplay between language contact and universals of language acquisition in particular. What has to be kept in mind when considering the findings of the present study is that in reality the language contact situations are, of course, much more complex, especially in multilingual societies. The present study can only give a simplified picture of the factors that interact in the shaping of linguistic knowledge in the different ecologies, especially in those where English is acquired as a second language. Recall that the Ethnologue lists 181 native languages for the Philippines, for example, and even 447 languages for India. Furthermore, recall that there are many Filipina domestic helpers in Hong Kong, who bring not only their variety of English but also Austronesian languages into the territory (Bolton 2003). It is, of course, not possible to investigate the information-packaging strategies in all these languages, although they might have an effect on the structuring of sentences in the respective English variety. Keeping this in mind, the present study finds different factors at work in the shaping of linguistic knowledge. For one, there are patterns of language use which are very likely due to universal acquisitional principles (e.g. left dislocation), while other patterns can clearly be accounted for in terms of influence from the background languages (e.g. existential *got* in Singapore English). Second, there are also many features which cannot be traced back to a single source and are very likely due to multiple causes (e.g. subject-zero relatives in *there*-existentials in Hong Kong English; frequency and variation of left dislocation in Irish English). And third, for yet other patterns of use a straightforward motivation is more difficult to find (e.g. the high frequency of left dislocation and fronting constructions in Indian English; small numbers of clefts in Singapore and Hong Kong English).

Hence, the present study confirms that the motivations for the emergence of certain linguistic structures are notoriously difficult to pin down (cf. Lange (2012: 150) on fronting constructions in Indian English). Yet still, the present study has contributed many new insights into the mechanisms of language contact and its interplay with language universals or the ecology. Some results call for further research though, especially research based on larger datasets. For example, right dislocation constructions that contain the particle *so* in the dislocated element (*so*-tags, e.g. *He's a real pet so he is*) seem to be a characteristic feature of Irish English, but the frequency of such tokens is so small in the present study and largely restricted to a single young

speaker and some elderly women that it is difficult to say whether it is a stable feature of the variety, old-fashioned or on the rise and used as an identity marker. Comments in the literature on this specific type of right dislocation are scarce, but it might be interesting to investigate the distribution of the construction in a large set of Irish English data in order to see whether we are here dealing with a pattern - well known from other studies - whereby a traditionally local feature is regaining ground in the youngest generation and thereby leading to a v-shaped age pattern (Dubois/Horvath 1999; Durham 2011).

A number of further questions are left unanswered. These include the distribution and use of 'hanging topic' constructions across varieties of English and their interaction with fronting and left dislocation constructions. The pragmatic functions of cleft constructions across varieties of English need to be studied more carefully, as well as those of existential *there*-constructions. Moreover, the low frequency of cleft constructions in Singapore English and Hong Kong English calls for further research. It has been suggested that the complexity of the cleft construction and influence from the Chinese L1s may be responsible for the low frequency, but these tentative suggestions clearly need back-up. Furthermore, throughout the study Jamaican English tends to pattern with the L1 English varieties as far as more complex structures are concerned. In other respects the variety behaves like the other L2 English varieties, however. It might be interesting to compare in some detail simple and complex variants of other structures and see whether Jamaican English behaves similarly in these cases. Especially rewarding might be a detailed analysis of the nature and use of relative clauses in Jamaican English and a comparison with other varieties of English. If Jamaican English speakers indeed used more complex structures more frequently than speakers of other L2 English varieties, the question of why they do not shy away from using such structures, as most other L2 English speakers tend to do, would be interesting to address, so it would.

CHAPTER 6

Appendix

6.1 ICE spoken texts

Hierarchical composition of the ICE spoken texts.[1]

Dialogue (180)	**S1**
Private (100)	**S1A**
Direct Conversations (90)	S1A-001 to S1A-090
Telephone Calls (10)	S1A-091 to S1A-100
Public (80)	**S1B**
Class Lessons (20)	S1B-001 to S1B-020
Broadcast Discussions (20)	S1B-021 to S1B-040
Broadcast Interviews (10)	S1B-041 to S1B-050
Parliamentary Debates (10)	S1B-051 to S1B-060
Legal Cross-examinations (10)	S1B-061 to S1B-070
Business Transactions (10)	S1B-071 to S1B-080
Monologue (120)	**S2**
Unscripted (70)	**S2A**
Spontaneous Commentaries (20)	S2A-001 to S2A-020
Unscripted Speeches (30)	S2A-021 to S2A-050
Demonstrations (10)	S2A-051 to S2A-060
Legal Presentations (10)	S2A-061 to S2A-070
Scripted (50)	**S2B**
Broadcast News (20)	S2B-001 to S2B-020
Broadcast Talks (20)	S2B-021 to S2B-040
Speeches (not Broadcast) (10)	S2B-041 to S2B-050

1 Source: The ICE project web page at http://ice-corpora.net/ice/design.htm.

6.2 ICE written texts

Hierarchical composition of the ICE written texts: categories and corresponding file names.[2]

Non-printed (50)	**W1**
Non-professional Writing (20)	**W1A**
Student Essays (10)	W1A-001 to W1A-010
Examination Scripts (10)	W1A-011 to W1A-020
Correspondence (30)	**W1B**
Social Letters (15)	W1B-001 to W1B-015
Business Letters (15)	W1B-016 to W1B-030
Printed (150)	**W2**
Academic Writing (40)	**W2A**
Humanities (10)	W2A-001 to W2A-010
Social Sciences (10)	W2A-011 to W2A-020
Natural Sciences (10)	W2A-021 to W2A-030
Technology (10)	W2A-031 to W2A-040
Non-academic Writing (40)	**W2B**
Humanities (10)	W2B-001 to W2B-010
Social Sciences (10)	W2B-011 to W2B-020
Natural Sciences (10)	W2B-021 to W2B-030
Technology (10)	W2B-031 to W2B-040
Reportage (20)	**W2C**
Press News Reports (20)	W2C-001 to W2C-020
Instructional Writing (20)	**W2D**
Administrative Writing (10)	W2D-001 to W2D-010
Skills & Hobbies (10)	W2D-011 to W2D-020
Persuasive Writing (10)	**W2E**
Press Editorials (10)	W2E-001 to W2E-010
Creative Writing (10)	**W2F**
Novels & Stories (10)	W2F-001 to W2F-020

2 Source: The ICE project web page at http://ice-corpora.net/ice/design.htm.

6.3 GloWbE

Composition of the Corpus of Global Web-based English.[3]

Country	Code	Web sites	Web pages	Words
United States	US	82,260	275,156	386,809,355
Canada	CA	33,776	135,692	134,765,381
Great Britain	GB	64,351	381,841	387,615,074
Ireland	IE	15,840	102,147	101,029,231
Australia	AU	28,881	129,244	148,208,169
New Zealand	NZ	14,053	82,679	81,390,476
India	IN	18,618	113,765	96,430,888
Sri Lanka	LK	4,208	38,389	46,583,115
Pakistan	PK	4,955	42,769	51,367,152
Bangladesh	BD	5,712	45,059	39,658,255
Singapore	SG	8,339	45,459	42,974,705
Malaysia	MY	8,966	45,601	42,420,168
Philippines	PH	10,224	46,342	43,250,093
Hong Kong	HK	8,740	43,936	40,450,291
South Africa	ZA	10,308	45,264	45,364,498
Nigeria	NG	4,516	37,285	42,646,098
Ghana	GH	3,616	47,351	38,768,231
Kenya	KE	5,193	45,962	41,069,085
Tanzania	TZ	4,575	41,356	35,169,042
Jamaica	JM	3,488	46,748	39,663,666
Total		**340,619**	**1,792,045**	**1,885,632,973**

3 Source: GloWbE web page at http://corpus.byu.edu/glowbe/.

6.4 COCA

Composition of the Corpus of Contemporary American English.[4]

Year	Spoken	Fiction	Magazine	Newspaper	Academic	Total
1990	4,332,983	4,176,786	4,061,059	4,072,572	3,943,968	20,587,368
1991	4,275,641	4,152,690	4,170,022	4,075,636	4,011,142	20,685,131
1992	4,493,738	3,862,984	4,359,784	4,060,218	3,988,593	20,765,317
1993	4,449,330	3,936,880	4,318,256	4,117,294	4,109,914	20,931,674
1994	4,416,223	4,128,691	4,360,184	4,116,061	4,008,481	21,029,640
1995	4,506,463	3,925,121	4,355,396	4,086,909	3,978,437	20,852,326
1996	4,060,792	3,938,742	4,348,339	4,062,397	4,070,075	20,480,345
1997	3,874,976	3,750,256	4,330,117	4,114,733	4,378,426	20,448,508
1998	4,424,874	3,754,334	4,353,187	4,096,829	4,070,949	20,700,173
1999	4,417,997	4,130,984	4,353,229	4,079,926	3,983,704	20,965,840
2000	4,414,772	3,925,331	4,353,049	4,034,817	4,053,691	20,781,660
2001	3,987,514	3,869,790	4,262,503	4,066,589	3,924,911	20,111,307
2002	4,329,856	3,745,852	4,279,955	4,085,554	4,014,495	20,455,712
2003	4,404,978	4,094,865	4,295,543	4,022,457	4,007,927	20,825,770
2004	4,330,018	4,076,462	4,300,735	4,084,584	3,974,453	20,766,252
2005	4,396,030	4,075,210	4,328,642	4,089,168	3,890,318	20,779,368
2006	4,304,513	4,081,287	4,279,043	4,085,757	4,028,620	20,779,220
2007	3,882,586	4,028,998	4,185,161	3,975,474	4,267,452	20,339,671
2008	3,635,622	4,155,298	4,205,477	4,031,769	4,015,545	20,043,711
2009	3,969,587	4,143,814	3,855,815	3,971,607	4,144,064	20,084,887
2010	4,095,393	3,929,160	3,806,011	4,258,633	3,816,420	19,905,617
2011	4,033,627	4,166,029	4,199,378	3,982,299	4,064,535	20,445,868
2012	2,348,159	2,294,570	2,203,821	2,109,683	2,298,658	11,254,891
Total	**95,385,672**	**90,344,134**	**95,564,706**	**91,680,966**	**91,044,778**	**464,020,256**

4 Source: COCA web page at http://corpus.byu.edu/coca/. The most recent addition of texts was completed in June 2012.

6.5 ICE word counts

Word counts of the 'direct conversation' files in ICE and the R code with which they have been computed.

Word counts of the S1A-files in ICE.

ICE-Great Britain	201,645
ICE-Ireland	201,695
ICE-New Zealand	230,007
ICE-Canada	210,671
ICE-Singapore	203,299
ICE-Philippines	216,609
ICE-Jamaica	213,279
ICE-India	216,010
ICE-Hong Kong	237,974

R code

```
sample <- readLines("C:/.../S1A-001.txt")                  # import text file into R
sample <-paste(ire, collapse="")                           # collapse all lines
sample <-gsub("<& >.*?</& >", "", sample)                  # remove annotations
sample <-gsub("<O>.*?</O>", "", sample)
sample <-gsub("<unclear>.*?</unclear>", "", sample)
sample <-gsub("<X>.*?</X>", "", sample)
sample <-gsub("<.*?>", "", sample)                         # remove tags
sample <-gsub("[[:punct:]]", "", sample)
trim <- function(x) {gsub("^\\s+| \\s+$", "", x)}          # remove whitespace
sample <-trim(sample)
sample <- gsub(" +", "", sample)                           # remove double whitespace
words <- strsplit(sample, "")                              # split into words
```

6.6 Left dislocation

6.6.1 Anaphoricity

Initial elements in LD constructions containing new, inferable or old information (token frequencies and percentages out of all LD tokens).

corpus	LD tokens	new		inferable		old	
		N	%	N	%	N	%
GB	72	27	37.5	17	23.6	28	38.9
IRE	129	51	39.5	26	20.2	52	40.3
NZ	95	33	34.7	22	23.2	40	42.1
CAN	97	40	41.2	21	21.6	36	37.1
SIN	117	41	35.0	22	18.8	54	46.2
PHI	169	39	23.1	26	15.4	104	61.5
JAM	169	51	30.2	45	26.6	73	43.2
IND	356	103	28.9	70	19.7	183	51.4
HK	170	50	29.4	36	21.2	84	49.4

6.6.2 Persistence

Persistence of the initial elements in LD constructions in the following discourse (token frequencies and percentages out of all LD tokens).

corpus	LD tokens	zero		NP/indirect		pronoun	
		N	%	N	%	N	%
GB	72	22	30.6	12	16.7	38	52.8
IRE	129	24	18.6	31	24.0	74	57.4
NZ	95	18	18.9	24	25.3	53	55.8
CAN	97	29	29.9	21	21.6	47	48.5
SIN	117	33	28.2	26	22.2	58	49.6
PHI	169	44	26.0	26	15.4	99	58.6
JAM	169	48	28.4	42	24.9	79	46.7
IND	356	101	28.4	100	28.1	155	43.5
HK	170	37	21.8	41	24.1	92	54.1

6.6.3 Complexity

LD tokens with a complex initial element (token frequencies, frequencies per 100,000 words and percentages out of all LD tokens).

corpus	complex		
	N	norm.	%
GB	23	11.4	31.9
IRE	27	13.4	20.9
NZ	22	23.2	23.4
CAN	28	13.3	28.9
SIN	12	5.9	10.3
PHI	20	9.2	11.8
JAM	46	21.6	27.2
IND	51	23.6	14.3
HK	16	6.7	9.4

6.6.4 *For*-LDs

***For*-LDs with pronominal and nominal preclausal elements in ICE and GloWbE (frequencies per 100,000 words).**

variety	ICE			GloWbE		
	pronoun	noun	total	pronoun	noun	total
GB	0.99	-	0.99	0.13	0.02	0.15
IRE	-	-	-	0.07	0.03	0.10
NZ	-	-	-	0.09	0.03	0.11
CAN	0.95	0.47	1.42	0.06	0.04	0.10
SIN	5.90	2.46	8.36	0.16	0.09	0.24
PHI	2.77	2.77	5.54	0.12	0.03	0.15
JAM	0.94	0.47	1.41	0.09	0.04	0.13
IND	2.31	1.39	3.70	0.05	0.02	0.06
HK	4.20	13.87	18.07	0.06	0.06	0.12

6.6.5 Pronominal LDs in GloWbE

Pronominal LDs with the initial sequence *me I* in GloWbE (frequencies per 100,000 words).

variety	LDs with *me I*
	norm.
GB	0.065
IRE	-
NZ	0.012
CAN	0.034
SIN	0.015
PHI	0.019
JAM	0.013
IND	0.009
HK	0.028

6.7 Right dislocation

6.7.1 RDs by gender

Canonical and expanded RDs by gender in the S1A-files of ICE-Ireland (token frequencies and percentages out of all RD tokens).

gender	ICE speakers		canonical RDs		expanded RDs	
	N	%	N	%	N	%
women	285	75.8	46	76.7	42	65.6
men	91	24.2	14	23.3	22	34.4

6.7.2 RDs by age group

Canonical and expanded RDs by age group in the S1A-files of ICE-Ireland (token frequencies and percentages out of all RD tokens).

age group	ICE speakers		RD tokens		canonical RDs		expanded RDs	
	N	%	N	%	N	%	N	%
19-25	163	52.1	50	49.0	21	44.7	29	52.7
26-33	69	22.0	17	16.7	9	19.1	8	14.5
34-41	15	4.8	6	5.9	1	2.1	5	9.1
42-49	9	2.9	2	2.0	2	4.3	-	-
50+	57	18.2	27	26.5	14	29.8	13	23.6

6.8 Fronting constructions

6.8.1 Anaphoricity

Initial elements in fronting constructions containing new, inferable or old information (token frequencies and percentages out of all fronting constructions).

corpus	FRON tokens	new		inferable		old	
		N	%	N	%	N	%
GB	35	7	20.0	11	31.4	17	48.6
IRE	51	11	21.6	16	31.4	24	47.1
NZ	36	8	22.2	19	52.8	9	25.0
CAN	28	10	35.7	7	25.0	11	39.3
SIN	53	8	15.1	20	37.7	25	47.2
PHI	25	4	16.0	5	20.0	16	64.0
JAM	24	3	12.5	12	50.0	9	37.5
IND	213	36	16.9	71	33.3	106	49.8
HK	14	-	-	7	50.0	7	50.0

6.8.2 Persistence

Persistence of the initial elements in fronting constructions in the following discourse (token frequencies and percentages out of all fronting constructions).

corpus	FRON tokens	zero		NP/indirect		pronoun	
		N	%	N	%	N	%
GB	35	17	48.6	9	25.7	9	25.7
IRE	51	23	45.1	16	31.4	12	23.5
NZ	36	14	38.9	16	44.4	6	16.7
CAN	28	5	17.9	17	60.7	6	21.4
SIN	53	25	47.2	20	37.7	8	15.1
PHI	25	10	40.0	10	40.0	5	20.0
JAM	24	11	45.8	11	45.8	2	8.3
IND	213	89	41.8	97	45.5	27	12.7
HK	14	9	64.3	3	21.4	2	14.3

6.8.3 Realization

Realization of the fronted element: noun phrases (nouns and pronouns), prepositional phrases, adjective phrases and clauses out of all fronting constructions.

corpus	NP		PP		AP		clause	
	N	%	N	%	N	%	N	%
GB	25	71.4	2	5.7	3	8.6	5	14.3
IRE	28	54.9	15	29.4	5	9.8	3	5.9
NZ	24	66.7	7	19.4	3	8.3	2	5.6
CAN	17	60.7	11	39.3	-	-	-	-
SIN	47	88.7	2	3.8	4	7.5	-	-
PHI	19	76.0	-	-	6	24.0	-	-
JAM	15	62.5	5	20.8	1	4.2	3	12.5
IND	159	74.6	33	15.5	18	8.5	3	1.4
HK	13	92.9	-	-	1	7.1	-	-

6.9 Existential *there*-constructions

6.9.1 Type of extension

Type of extension in complex existential *there*-constructions (token frequencies and percentages out of all extended existentials).

corpus	adverbial		predicative		infinitival		participial		relclause	
	N	%	N	%	N	%	N	%	N	%
GB	138	37.5	12	3.3	14	3.8	67	18.2	137	37.2
IRE	222	51.3	14	3.2	20	4.6	88	20.3	89	20.6
NZ	145	41.2	15	4.3	30	8.5	56	15.9	106	30.1
CAN	133	42.4	9	2.9	21	6.7	37	11.8	114	36.3
SIN	62	32.0	5	2.6	23	11.9	33	17.0	71	36.6
PHI	103	38.3	12	4.5	14	5.2	35	13.0	105	39.0
JAM	59	20.1	14	4.8	29	9.9	23	7.8	168	57.3
IND	130	48.9	11	4.1	19	7.1	21	7.9	85	32.0
HK	130	42.6	11	3.6	24	7.9	40	13.1	100	32.8

6.9.2 Singular concord: overall distribution

Number of existential *there*-constructions with plural notional subjects; token frequencies and proportions of singular concord out of these.

corpus	plural	singular concord	
		N	%
GB	207	67	32.4
IRE	229	168	73.4
NZ	211	145	68.7
CAN	135	76	56.3
SIN	141	31	22.0
PHI	189	29	15.3
JAM	133	14	10.5
IND	201	19	9.5
HK	268	67	25.0

6.9.3 Singular concord: influencing factors

Factors influencing concord variation (raw frequencies and percentages out of all *there*-existentials with plural notional subjects).[5]

corpus	concord	total		present		past		contracted		bare		negated	
		N	%	N	%	N	%	N	%	N	%	N	%
GB	plural	140	67.6	95	67.9	41	29.3	11	7.9	56	40.0	7	5.0
	singular	67	32.4	65	97.0	2	3.0	62	92.5	29	43.3	5	7.5
IRE	plural	61	26.6	34	55.8	27	44.3	2	3.3	29	47.5	4	6.6
	singular	168	73.4	119	70.8	49	29.2	106	63.1	72	42.9	13	7.7
NZ	plural	66	31.3	36	54.5	30	45.5	3	4.5	25	37.9	12	18.2
	singular	145	68.7	100	69.0	44	30.3	98	67.6	66	45.5	21	14.5
CAN	plural	59	43.7	34	57.6	24	40.7	10	17.0	24	40.7	4	6.8
	singular	76	56.3	71	93.4	5	6.6	69	90.8	42	55.3	6	8.0
SIN	plural	110	78.0	96	87.3	14	12.7	15	13.6	57	51.8	7	6.3
	singular	31	22.0	28	90.3	3	9.7	20	64.5	14	45.2	3	9.7
PHI	plural	160	84.7	113	70.6	47	29.4	6	3.8	74	46.3	8	5.0
	singular	29	15.3	26	89.7	3	10.3	19	65.5	13	44.8	2	6.9
JAM	plural	119	89.5	107	89.9	12	10.1	27	22.7	35	29.4	4	3.4
	singular	14	10.5	13	92.9	1	7.1	8	57.1	7	50.0	6	40.0
IND	plural	182	90.6	169	92.9	13	7.1	10	5.5	84	46.2	15	8.1
	singular	19	9.5	15	79.0	4	21.1	4	21.1	15	79.0	3	15.0
HK	plural	201	75.0	187	92.5	15	7.5	27	13.4	92	45.8	15	7.5
	singular	67	25.0	67	100	-	-	52	77.6	31	46.3	10	14.9

5 The percentages are to be understood as follows: in the British English data, for example, the category 'bare' says that in the case of plural concord 40% are bare existentials and 60% are extended existentials.

6.9.4 Singular concord: determiner type

Determiner types in *there*-existentials (raw frequencies and percentages out of all existentials with plural notional subjects).[6]

corpus	concord	bare		definite		number		negative		adjective		quantifier		a	
		N	%	N	%	N	%	N	%	N	%	N	%	N	%
GB	plural	27	19.3	6	4.3	28	20.0	5	3.6	21	15.0	45	32.1	8	5.7
	singular	10	14.9	5	7.5	16	23.9	4	6.0	2	3.0	20	29.9	10	14.9
IRE	plural	13	21.3	-	-	12	19.7	3	4.9	9	14.8	19	31.1	4	6.6
	singular	24	14.2	1	0.6	57	33.7	13	7.7	16	9.5	42	24.9	14	8.3
NZ	plural	10	15.2	4	6.1	14	21.2	4	6.1	6	9.1	19	28.8	9	13.6
	singular	27	18.6	4	2.8	30	20.7	16	11.0	17	11.7	36	24.8	15	10.3
CAN	plural	16	27.1	1	1.7	14	23.7	1	1.7	5	8.5	17	28.8	5	8.5
	singular	14	18.7	4	5.3	18	24.0	6	8.0	6	8.0	14	18.7	13	17.3
SIN	plural	16	14.3	-	-	15	13.4	7	6.3	13	11.6	39	34.8	22	19.6
	singular	5	16.1	3	9.7	6	19.4	3	9.7	3	9.7	9	29.0	2	6.5
PHI	plural	51	31.5	3	1.9	21	13.0	5	3.1	17	10.5	45	27.8	19	11.7
	singular	6	20.7	2	6.9	6	20.7	1	3.4	1	3.4	7	24.1	6	20.7
JAM	plural	40	33.6	2	1.7	11	9.2	4	3.4	15	12.6	30	25.2	17	14.3
	singular	-	-	1	6.7	2	13.3	6	40.0	1	6.7	2	13.3	3	20.0
IND	plural	26	14.1	2	1.1	31	16.8	15	8.1	20	10.8	84	45.4	7	3.8
	singular	2	10.0	1	5.0	3	15.0	3	15.0	1	5.0	6	30.0	4	20.0
HK	plural	19	9.4	2	1.0	28	13.8	14	6.9	24	11.8	111	54.7	5	2.5
	singular	3	4.5	1	1.5	11	16.4	10	14.9	6	9.0	32	47.8	4	6.0

6 The percentages are to be understood as follows: in the Irish English data, for example, 21.3% of existentials with plural concord are bare existentials, while 19.7% of existentials with plural concord have a number before the notional subject.

6.10 Cleft constructions

6.10.1 Subordinator types in it-clefts

Distribution of subordinator types in IT-clefts: *wh*-words, *that*, *ing*-complements, zero and other minor types (token frequencies and percentages out of all IT-clefts).

corpus	wh		that		zero		ing		other*	
	N	%	N	%	N	%	N	%	N	%
GB	14	23.3	27	45.0	16	26.7	1	1.7	2	3.3
IRE	6	5.5	23	21.1	59	54.1	10	9.2	11	10.1
NZ	5	9.4	25	47.2	12	22.6	9	17.0	2	3.8
CAN	7	14.6	27	56.3	12	25.0	-	-	2	4.2
SIN	6	19.4	14	45.2	7	22.6	-	-	4	12.9
PHI	16	32.0	18	36.0	5	10.0	2	4.0	9	18.0
JAM	13	18.1	37	51.4	20	27.8	1	1.4	1	1.4
IND	9	30.0	9	30.0	11	36.7	-	-	1	3.3
HK	4	16.7	9	37.5	8	33.3	1	4.2	2	8.3

*The category 'other' comprises incomplete constructions and clefts with the subordinators *since*, *till*, *for* and *and*.

6.10.2 Types of basic pseudo-cleft

Distribution of basic pseudo-clefts introduced by a *wh*-word, *all* or a pro-noun (token frequencies and percentages out of all basic pseudo-clefts).

corpus	wh-word		all		pro-noun	
	N	%	N	%	N	%
GB	77	62.1	16	12.9	31	25.0
IRE	26	36.6	22	31.0	23	32.4
NZ	57	60.0	19	20.0	19	20.0
CAN	78	61.9	19	15.1	29	23.0
SIN	73	75.3	7	7.2	17	17.5
PHI	103	77.4	10	7.5	20	15.0
JAM	109	66.9	21	12.9	33	20.2
IND	127	90.1	2	1.4	12	8.5
HK	29	53.7	10	18.5	15	27.8

6.10.3 Types of reversed pseudo-cleft

Distribution of reversed pseudo-clefts whose cleft clause begins with a *wh*-word, *all* or a pro-noun (token frequencies and percentages out of all reversed pseudo-clefts).

corpus	wh-word		all		pro-noun	
	N	%	N	%	N	%
GB	179	73.4	12	4.9	53	21.7
IRE	198	73.3	18	6.7	54	20.0
NZ	233	77.7	16	5.3	51	17.0
CAN	224	75.4	9	3.0	64	21.5
SIN	183	82.1	6	2.7	34	15.2
PHI	214	69.7	7	2.3	86	28.0
JAM	220	74.3	6	2.0	70	23.6
IND	183	88.4	2	1.0	22	10.6
HK	99	85.3	6	5.2	11	9.5

References

Aijmer, Karin. 1989. Themes and Tails: The Discourse Function of Dislocated Elements. *Nordic Journal of Linguistics* 12 (2): 137-154.

Akmajian, Adrian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistic Inquiry* 1 (2): 149-168.

Allerton, D. J. 1991. The greater precision of spoken language: four examples from English. *English Studies* 5: 470-478.

Alleyne, Mervyn C. 1984. *Roots of Jamaican culture*. London: Pluto.

Alsagoff, Lubna. 2010. English in Singapore: culture, capital and identity in linguistic variation. World Englishes 29 (3): 336–348.

Alsagoff, Lubna and Ho Chee Lick. 1998. The grammar of Singapore English. In Foley, ed.: 127-151.

Ansaldo, Umberto. 2009. The Asian typology of English: theoretical and methodological considerations. *English World-Wide* 30 (2): 133-148.

Asher, R. E. and T. C. Kumari. 1997. *Malayalam*. London/New York: Routledge.

Baayen, Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bailey, Beryl L. 1966. *Jamaican Creole syntax: a transformational approach*. Cambridge: Cambridge University Press.

Bailey, Beryl L. 1971. Jamaican Creole: can dialect boundaries be defined? In Hymes, ed.: 341-348.

Ball, Catherine N. 1994. The origins of the informative-presupposition *it*-cleft. *Journal of Pragmatics* 22: 603-628.

Bao, Zhiming. 2005. The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics* 41 (2): 237–67.

Bao, Zhiming. 2010. A usage-based approach to substratum transfer: the case of four unproductive features in Singapore English. *Language* 86 (4): 792-820.

Bao, Zhiming. 2014. *Got* in Singapore English. In Eugene Green and Charles F. Meyer, eds. *The variability of current World Englishes*. Berlin: De Gruyter. 147-165.

Bauer, Laurie. 1994. English in New Zealand. In Burchfield, ed.: 382-429.

Bauer, Laurie. 2000. The dialectal origins of New Zealand English. In Bell and Kuiper, eds.: 40-52.

Bauer, Robert S. and Stephen Matthews. 2003. Cantonese. In Graham Thurgood and Randy J. LaPolla, eds. *The Sino-Tibetan languages*. London: Rouledge. 146-155.

Bautista, Ma. Lourdes S. and Kingsley Bolton, eds. 2008. *Philippine English: linguistic and literary perspectives*. Hong Kong: Hong Kong University Press.

Beal, Kalynda. 2012. Is it truly unique that Irish English clefts are? Quantifying the syntactic variation of it-clefts in Irish English and other post-colonial English varieties. In Bettina Migge, ed. *New perspectives on Irish English.* Amsterdam: John Benjamins. 153-177.

Beckford Wassink, Alicia. 1999. Historic low prestige and seeds of change: attitudes toward Jamaican Creole. *Language in Society* 28 (1): 57-92.

Bell, Allan. 1997. The phonetics of fish and chips in New Zealand: marking national and ethnic identities. *English World-Wide* 18 (2): 243–270.

Bell, Allan. 2000. Maori and Pakeha English: a case study. In Bell and Kuiper, eds.: 221-248.

Bell, Allan and Koenraad Kuiper, eds. 2000. *New Zealand English.* Amsterdam: John Benjamins.

Benton, Richard A. 1991. Maori English: a New Zealand myth? In Jenny Cheshire, ed. *English around the world: sociolinguistic perspectives.* Cambridge: Cambridge University Press. 187-199.

Bergs, Alexander and Laurel J. Brinton. 2012. *English historical linguistics: an international handbook.* Berlin: De Gruyter.

Bhatt, Rakesh M. 2004. Indian English: syntax. In Kortmann and Schneider, eds.: 1016-1030.

Bhatt, Rakesh M. 2008. Indian English: syntax. In Rajend Mesthrie, ed. *Varieties of English 4: Africa, South and Southeast Asia.* 546-562.

Biber, Douglas, et al. 1999. *Longman grammar of spoken and written English.* Harlow: Longman.

Bickerton, Derek. 1973. The nature of the creole continuum. *Language* 49 (3): 640-669.

Bickerton, Derek. 1975. *Dynamics of a creole system.* Cambridge: Cambridge University Press.

Birner, Betty J. 1997. The linguistic realization of inferrable information. *Language & Communication* 17 (2): 133-147.

Birner, Betty J. and Gregory L. Ward. 1998. *Information status and noncanonical word order in English.* Amsterdam: John Benjamins.

Bliss, A. 1984. English in the south of Ireland. In Peter Trudgill, ed. *Language in the British Isles.* Cambridge: Cambridge University Press. 135-151.

Bloomfield, Morton. 1948. Canadian English and its relation to eighteenth century American speech. *Journal of English and Germanic Philology* 47 (1): 59–67.

Boberg, Charles. 2004. The dialect topography of Montreal. *English World-Wide* 25 (2): 171–198.

Boberg, Charles. 2010. *The English language in Canada : status, history and comparative analysis.* Cambridge: Cambridge University Press.

Bolinger, Dwight. 1977. *Meaning and form.* London: Longman.

Bolton, Kingsley. 2000. The sociolinguistics of Hong Kong and the space for Hong Kong English. *World Englishes* 19 (3): 265-285.

Bolton, Kingsley. 2003. *Chinese Englishes: a sociolinguistic history.* Cambridge: Cambridge University Press.

Bolton, Kingsley and Ma. Lourdes S. Bautista. 2008. Philippine English: linguistic and literary perspectives. In Bautista and Bolton, eds.: 1-9.

Borlongan, Ariane Macalinga. 2011. Relocating Philippine English in Schneider's dynamic model. Paper presented at the 17th Annual Conference of the International Association for World Englishes, Melbourne, Australia.

Breivik, Leiv Egil. 1981. On the interpretation of existential *there*. *Language* 57 (1): 1-25.

Breivik, Leiv Egil. 1990. *Existential* there*: a synchronic and diachronic study*. Oslo: Novus Press.

Breivik, Leiv Egil. 1999. On the rhetorical function of existential *there*. *Nordlit* 6: 3-14. http://www.hum.uit.no/nordlit/6/breivik.html (accessed 18 September 2014).

Breivik, Leiv Egil and Toril Swan. 2000. The desemanticisation of existential *there* in a synchronic-diachronic perspective. In Christiane Dalton-Puffer and Nikolaus Ritt, eds. *Words: structure, meaning, function: a festschrift for Dieter Kastovsky*. Berlin: De Gruyter. 19-33.

Breivik, Leiv Egil and Ana E. Martinez Insua. 2008. Grammaticalization, subjectification and non-concord in English existential sentences. *English Studies* 89 (3): 351-362.

Britain, David. 1992. Linguistic change in intonation: the use of High Rising Terminals in New Zealand English. *Language Variation and Change* 4 (1): 77–104.

Britain, David. 2000. *As far as* analysing grammatical variation and change in New Zealand English with very few tokens <*is concerned/∅*>. In Bell and Kuiper, eds.: 198-220.

Britain, David and Andrea Sudbury. 2002. "There's sheep and there's penguins": convergence, 'drift' and 'slant' in New Zealand and Falkland Island English. In M.C. Jones and E. Esch, eds. *Language change: the interplay of internal, external and extra-linguistic factors*. Berlin: De Gruyter. 211–240.

Brown, Keith, ed. 2006. Encyclopedia of language and linguistics (2. ed.). Amsterdam: Elsevier. http://www.sciencedirect.com/science/referenceworks/9780080448541 (accessed 8 July 2014).

Budge, Carol. 1989. Plural marking in Hong Kong English. *Hong Kong Papers in Linguistics and Language Teaching* 12. 39-47.

Burchfield, Robert, ed. 1994. *The Cambridge history of the English language*. Cambridge: Cambridge University Press.

Buschfeld, Sarah, Thomas Hoffmann, Magnus Huber and Alexander Kautzsch, eds. 2014. *The evolution of Englishes: the Dynamic Model and beyond*. Amsterdam: John Benjamins.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Bybee, Joan L. and Paul Hopper, eds. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, Joan L., Revere D. Perkins and William Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Calude, Andreea. 2008. Demonstrative clefts and double cleft constructions in spontaneous spoken English. *Studia Linguistica* 62 (1): 78-118.

Calude, Andreea. 2009a. *Cleft Constructions in Spoken English*. Saarbrücken: VDM

Verlag Dr. Müller.

Calude, Andreea. 2009b. Clefting and extraposition in English. *ICAME Journal* No.32. http://clu.uni.no/icame/ij32/index.html (accessed 19 August 2014).

Calude, Andreea and Steven Miller. 2009. Are clefts contagious in conversation? *English Language and Linguistics* 13 (1): 127-132.

Carter, Ronald and Michael McCarthy. 1995. Grammar and the spoken language. *Applied Linguistics* 16 (2): 141-158.

Carter, Ronald and Michael McCarthy. 2006. *Cambridge Grammar of English: a comprehensive guide.* Cambridge: Cambridge University Press.

Carter, Ronald, Rebecca Hughes and Michael McCarthy. 1998. Telling tails: grammar, the spoken language and materials development. In Brian Tomlinson, ed. *Materials development in language teaching.* Cambridge: Cambridge University Press. 78-100.

Cassidy, Frederic G. and Robert LePage. 1980. *Dictionary of Jamaican English.* Cambridge: Cambridge University Press.

Central Statistics Office. 2014. *Census 2011: Ireland and Northern Ireland.* http://www.cso.ie/en/census/index.html (accessed 22 October 2014).

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, ed.: 25-55.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time.* Chicago: University of Chicago Press.

Chambers, Jack K. 1973. Remarks on Topicalization in Child Language. *Foundations of Language* 9 (3): 442–446.

Chambers, Jack K. 1998. English: Canadian varieties. In John Edwards, ed. *Language in Canada.* Cambridge: Cambridge University Press. 252–272.

Chambers, Jack K. 2006. The development of Canadian English. In Kingsley Bolton and Braj B. Kachru, eds. *World Englishes: Critical Concepts in Linguistics.* London: Routledge. 383–395.

Chan, Alice Y.W. 2004. Syntactic transfer: evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal* 88: 56–74.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese.* Berkeley, Calif.: University of California Press.

Cheng, Lisa Lai-Shen. 2008. Deconstructing the *shì…de* construction. *The Linguistic Review* 25 (3/4): 235-266.

Chomsky, Noam. 1971. Deep structure, surface structure, and semantic interpretation. In Danny D. Steinberg and Leon A. Jakobovits, eds. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology.* Cambridge: Cambrdige University Press. 183–216.

Christie, Pauline. 1996. Jamaican relatives in focus. In Pauline Christie, ed. *Caribbean language issues: old and new: papers in honour of Professor Mervyn Alleyne on the occasion of his sixtieth birthday.* Kingston, Jamaica: The Press University of the West Indies. 48-60.

Cinque, Guglielmo. 1993. A Null Theory of Phrase and Compound Stress. *Linguistic Inquiry* 24 (2): 239-297.

Clarke, Sandra. 2010. *Newfoundland and Labrador English.* Edinburgh: Edinburgh University Press.

Collins, Peter C. 1991. *Cleft and pseudo-cleft constructions in English.* London: Routledge.
Collins, Peter C. 2006. *It*-clefts and *wh*-clefts: prosody and pragmatics. *Journal of Pragmatics* 38: 1706-1720.
Collins, Peter C. 2012. Singular agreement in *there*-existentials: an intervarietal corpus-based study. *English World-Wide* 33 (1): 53-68.
Collins, Peter C. and Pam Peters. 2004. Australian English: morphology and syntax. In Kortmann and Schneider, eds.: 593-642.
Collins, Peter C., Ariane Macalinga Borlongan and Xinyue Yao. 2014. Modality in Philippine English: a diachronic study. *Journal of English Linguistics* 42 (1): 68-88.
Constitution of Ireland. 2013. http://www.taoiseach.gov.ie/eng/Historical_Information/The_Constitution/ (accessed 23 October 2014).
Cotton, Eleanor G. 1978. Noun-pronoun pleonasms: the role of age and situation. *Journal of Child Language* 5 (3): 489-499.
Crawford, William J. 2005. Verb agreement and disagreement: a corpus investigation of concord variation in existential *there* + *be* constructions. *Journal of English Linguistics* 33 (1): 35–61.
Crowley, Tony. 2000. *The politics of language in Ireland, 1366-1922: a sourcebook.* London: Routledge.
Crystal, David. 2003. *English as a global language.* Cambridge: Cambridge University Press.
Cullen, Richard and I-Chun (Vicky) Kuo. 2007. Spoken grammar and ELT course materials: a missing link? *TESOL Quarterly* 41 (2): 361-386.
Danielewicz-Betz, Anna and David Graddol. 2014. Varieties of English in the urban landscapes of Hong Kong and Shenzhen. *English Today* 30 (3): 22-32.
Davidse, Kristin. 2000. A constructional approach to clefts. *Linguistics* 38 (6): 1101–31.
DeCamp, David. 1971. Towards a generative analysis of a post-creole speech continuum. In Hymes, ed.: 149-370.
Declerck, Renaat. 1983. Predicational clefts. *Lingua* 61: 9-45.
Declerck, Renaat. 1984. The pragmatics of *it*-clefts and *wh*-clefts. *Lingua* 64: 251-289.
Declerck, Renaat. 1992. The inferential *it is that*-construction and its congeners. *Lingua* 87: 203-230.
De Fréine, Séamus. 1977. The dominance of the English language in the 19th century. In Diarmuid Ó Muirithe, ed. *The English Language in Ireland.* Cork: Mercier. 127–49.
Dehé, Nicole and Anne Wichmann. 2010. Sentence-initial *I think (that)* and *I believe (that)*: prosodic evidence for use as main clause, comment clause and discourse marker. *Studies in Language* 34 (1): 36–74.
Delahunty, Gerald P. 1995. The inferential construction. *Pragmatics* 5 (3): 341-364.
Delin, Judy and Jon Oberlander. 1995. Syntactic constraints on discourse structure: the case of *it*-clefts. *Linguistics* 33: 465-500.
Del Gobbo, Francesca. 2014. Classifiers. In Huang, Li and Simpson, eds.: 26-48.
Denison, David. 1999. Syntax. In Suzanne Romaine, ed. *The Cambridge History of the English Language Volume 4: 1776–1997.* Cambridge: Cambridge University Press. 92-329.

Department of Statistics Singapore. 2014. Census of population 2010, Statistical Release 1: Demographic Characteristics, Education, Language and Religion. http://www.singstat.gov.sg/ (accessed 5 August 2014).

De Smet, Hendrik. 2013. *Spreading patterns: diffusional change in the English system of complementation.* New York: Oxford University Press.

Deterding, David. 2007. *Singapore English* (Dialects of English). Edinburgh: Edinburgh University Press.

Deuber, Dagmar. 2009. Caribbean ICE corpora: some issues for fieldwork and analysis. In Andreas H. Jucker, Daniel Schreier and Marianne Hundt, eds. *Corpora: pragmatics and discourse.* Amsterdam: Rodopi. 425-450.

Deuber, Dagmar. 2014. *English in the Caribbean. Variation, style and standards in Jamaica and Trinidad.* Cambridge: Cambridge University Press.

Devonish, Hubert. 1998. On the existence of autonomous language varieties in 'creole continuum situations'. In Pauline Christie, Barbara Lalla, Velma Pollard and Lawrence Carrington, eds. *Studies in Caribbean language II.* St. Augustine: Society for Caribbean Linguistics. 1-12.

Devonish, Hubert. 2003. Language advocacy and 'conquest' diglossia in the 'anglophone' Caribbean. In Christian Mair, ed. *The politics of English as a world language: new horizons in postcolonial cultural studies.* Amsterdam: Rodopi. 157-177.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25 (2): 108–127.

Di Tullio, Á. 2006. Clefting in spoken discourse. In Brown, ed.: 483-491.

Dollinger, Stefan. 2008. *New-Dialect Formation in Canada: Evidence from the English Modal Auxiliaries.* Amsterdam: John Benjamins.

Dollinger, Stefan. 2012. Varieties of English: Canadian English in real-time perspective. In Bergs and Brinton, eds.: 1858-1879.

Drubig, H. B. and W. Schaffar. 2001. Focus constructions. In Haspelmath et al., eds.: 1079-1104.

Dubois, Sylvie and Barbara Horvath. 1999. When the music changes, you change too: gender and language change in Cajun English. *Language Variation and Change* 11 (3): 287–313.

Duffley, Patrick J. 2000. Gerund versus infinitive as complement of transitive verbs in English: the problems of "tense" and "control". *Journal of English Llinguistics* 28 (3): 221-248.

Durham, Mercedes. 2011. Right dislocation in Northern England: frequency and use — perception meets reality. *English World-Wide* 32 (3): 257–279.

Eisikovits, Edina. 1991. Variation in subject-verb agreement in Inner Sydney English. In Jenny Cheshire, ed. *English Around the World.* Cambridge: Cambridge University Press. 235–55.

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74 (2): 245-273.

É. Kiss, Katalin. 2001. Discourse configurationality. In Haspelmath et al., eds.: 1442-1455.

Erteschik-Shir, Nomi. 2007. *Information structure. The syntax-discourse interface.* Oxford: Oxford University Press.

Fanego, Teresa. 1996. The development of gerunds as objects of subject-control verbs in English (1400-1760). *Diachronica* 13 (1): 29-62.

Fanego, Teresa. 2007. Drift and the development of sentential complements in British and American English from 1700 to the present day. In Javier Pérez-Guerra, Dolores González-Álvarez, Jorge L. Bueno-Alonso, Esperanza Rama-Martínez, eds. *'Of varying language and opposing creed': New insights into Late Modern English.* Bern: Peter Lang. 161-235.

Ferguson, Charles A. 1959. Diglossia. *Word* 15: 325-340.

Filppula, Markku. 1999. *The grammar of Irish English. Language in Hibernian style.* London: Routledge.

Filppula, Markku. 2006. The making of Hiberno-English and other "Celtic Englishes". In Van Kemenade and Los, eds.: 507-536.

Filppula, Markku. 2009. The rise of *it*-clefting in English: areal-typological and contact-linguistic considerations. *English Language and Linguistics* 13: 267-293.

Filppula, Markku. 2012. Irish English. In Kortmann and Lunkenheimer, eds.: 30-47.

Filppula, Markku and Juhani Klemola. 2012. English in contact: Celtic and Celtic Englishes. In Bergs and Brinton, eds.: 1687-1703.

Firbas, Jan. 1964. On defining the theme in functional sentence analysis. *Travaux Linguistiques de Prague* 1: 267-280.

Fischer, Olga. 1992. Syntax. In Norman Blake, ed. *The Cambridge History of the English Language Volume 2: 1066–1476.* Cambridge: Cambridge University Press. 207-408.

Foley, Joseph. 1988. *New Englishes. The case of Singapore.* Singapore: Singapore University Press.

Foley, Joseph, ed. 1998. *English in new cultural contexts: reflections from Singapore.* Singapore: Oxford University Press.

Freeze, Ray. 2001. Existential constructions. In Haspelmath et al., eds.: 941-953.

Fuchs, Robert. 2012. Focus marking and semantic transfer in Indian English: The case of *also*. *English World-Wide* 33 (1): 27-53.

Geluykens, Ronald. 1988. Five types of clefting in English discourse. *Linguistics* 26: 823-841.

Geluykens, Ronald. 1992. *From discourse process to grammatical construction of left-dislocation in English.* Amsterdam: John Benjamins.

Gisborne, Nikolas. 2009. Aspects of the morphosyntactic typology of Hong Kong English. *English World-Wide* 30 (2): 149–169.

Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. In Li, ed.: 149-188.

Givón, Talmy. 1983. *Topic continuity in discourse: a quantitative cross-language study.* Amsterdam: John Benjamins.

Görlach, Manfred. 1991. *Englishes: studies in varieties of English 1984-1988.* Amsterdam: John Benjamins.

Gonzalez, Andrew. 1997. The history of English in the Philippines. In Ma. Lourdes S. Bautista, ed. *English Is an Asian language: the Philippine context.* Manila: The Macquarie Library. 25-40.

Gonzalez, Andrew. 2008. A favorable climate and soil: a transplanted language and literature. In Bautista and Bolton, eds.: 13-27.

Gordon, Elizabeth and Tony Deverson. 1998. *New Zealand English and English in New Zealand*. Auckland: New House.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury and Peter Trudgill, eds. 2004. *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.

Government of India, Ministry of Home Affairs, Office of the Registrar General & Census Commissioner. 2001. 2001 Census of India. http://www.censusindia.gov.in (accessed 3 September 2014).

Graddol, David. 2013. *Profiling English in China: The Pearl River Delta*. Cambridge: Cambridge English Language Assessment. http://www.cambridgeenglish.org (accessed 11 September 2014).

Greenbaum, Sidney, ed. 1996. *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Greenbaum, Sidney. 1996a. Introducing ICE. In Greenbaum, ed.: 3-12.

Greenblatt, Stephen, Carol T. Christ and Catherine Robson, eds. 2006. *The Norton anthology of English literature - The Victorian age*. New York: Norton.

Gregory, Michelle L. and Laura A. Michaelis. 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics* 33: 1665–1706.

Grice, Paul. 1989. *Studies in the way of words*. Cambridge, Mass.: Harvard University Press.

Grosz, Barbara and Yael Ziv. 1998. Centering, global focus, and right dislocation. In Marilyn Walker, Aravind Joshi and Ellen Prince, eds. *Centering Theory in Discourse*. Oxford: Clarendon. 293–308.

Groves, Julie. 2009. Hong Kong English - Does it exist? *HKBU Papers in Applied Language Studies* 13: 54-79.

Gruber, Jeffrey S. 1967. Topicalization in Child Language. *Foundations of Language* 3 (1): 37–65.

Grzegorek, Maria. 1984. English sentences with introductory *there* and their Polish counterparts. In Jacek Fisiak, ed. *Contrastive linguistics: prospects and problems*. Berlin: Mouton. 143-169.

Gundel, Jeanette K. 1977. Where do cleft sentences come from? *Language* 53 (3): 543-559.

Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith Moravcsik and Jessica Wirth, eds. *Studies in syntactic typology*. Amsterdam: John Benjamins. 209-239.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69: 274–307.

Gundel, Jeanette K. and Thorstein Fretheim. 2004. Topic and focus. In Laurence R. Horn and Gregory Ward, eds. *The handbook of pragmatics*. Malden, MA: Blackwell. 175-196.

Gupta, Anthea Fraser. 1998. The situation of English in Singapore. In Joseph A. Foley et al., eds.: 106–126.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English: part 2. *Journal of Linguistics* 3 (2): 199-244.

Hannay, Michael. 1985. *English existentials in functional grammar*. Dordrecht: Foris Publications.

Harris, John. 1993. The grammar of Irish English. In James Milroy and Lesley Milroy, eds. *Real English. The grammar of English dialects in the British Isles.* London: Longman. 139-186

Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible, eds. 2001. *Language Typology and Language Universals* Band 2. Berlin: De Gruyter.

Haugland, Kari E. 1993. A note on cleft and existential sentences in Old English. *English Studies: A Journal of English Language and Literature* 74 (5): 407-413.

Hay, Jennifer and Daniel Schreier. 2004. Reversing the trajectory of language change: subject-verb agreement with *be* in New Zealand English. *Language Variation and Change* 16: 209-235.

Hay, Jennifer, Margaret Maclagan and Elizabeth Gordon. 2008. *New Zealand English.* Edinburgh: Edinburgh University Press.

Heine, Bernd. 2008. Contact-induced word order change without word order change. In Siemund and Kintana, eds.: 33-60.

Heine, Bernd, Ulrike Claudi and Friederike Hünnemeyer. 1991. *Grammaticalization: a conceptual framework.* Chicago: University of Chicago Press.

Heine, Bernd and Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27 (3): 529-572.

Heine, Bernd and Tania Kuteva. 2005. *Language contact and grammatical change.* Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva. 2010. Contact and grammaticalization. In Hickey, ed.: 86-105.

Her, One-Soon. 1991. On the Mandarin possessive and existential verb *you* and its idiomatic expressions. *Language Sciences* 13 (3/4): 381-398.

Her, One-Soon. 2012. Distinguishing classifiers and measure words: a mathematical perspective and implications. *Lingua* 122 (14): 1668-1691.

Her, One-Soon and Chen-Tien Hsieh. 2010. On the semantic distinction between classifiers and measure words in Chinese. *Language and Linguistics* 11 (3): 527-551.

Herriman, Jennifer. 2004. Identifying relations: the semantic functions of *wh*-clefts in English. *Text* 24 (4): 447-469.

Hickey, Raymond. 2005. *Dublin English: evolution and change.* Amsterdam: John Benjamins.

Hickey, Raymond. 2007. *Irish English. History and present-day forms.* Cambridge: Cambridge University Press.

Hickey, Raymond, ed. 2010. *The handbook of language contact.* Chichester: Wiley-Blackwell.

Hickey, Raymond, ed. 2012. *Areal features of the anglophone world.* Berlin: De Gruyter.

Hickey, Raymond. 2012a. English in Ireland. In Hickey, ed.: 79-107.

Himmelmann, Nikolaus P. 2005. Tagalog. In Alexander Adelaar and Nikolaus P. Himmelmann, eds. *The Austronesain languages of Asia and Madagascar.* London: Routledge. 350-376.

Hinrichs, Lars. 2006. *Codeswitching on the Web: English and Jamaican Creole in e-mail communication.* Amsterdam: John Benjamins.

Hockett, Charles. 1958. *A course in modern linguistics.* New York: Macmillan.

Hole, Daniel. 2011. The deconstruction of Chinese *shì…de* clefts revisited. Lingua 121: 1707-1733.

Holm, John A. 1988. *Pidgins and creoles, vol. 1: theory and structure.* Cambridge: Cambridge University Press.

Holm, John A. 1989. *Pidgins and creoles, vol.2: reference survey.* Cambridge: Cambridge University Press.

Holm, John A. 1994. English in the Caribbean. In Burchfield, ed.: 328-381.

Holmes, Janet. 1996. The New Zealand spoken component of ICE: some methodological challenges. In Greenbaum, ed.: 163-178.

Hopper, Paul J. and Elizabeth Cross Traugott. 2003. *Grammaticalization.* Cambridge: Cambridge University Press.

Hornby, Peter A. 1971. Surface structure and the topic-comment distinction: a developmental study. *Child Development* 42 (6): 1975-1988.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15 (4): 531-564.

Huang, C.-T. James, Y.-H. Audrey Li and Andrew Simpson, eds. 2014. *The handbook of Chinese linguistics.* Malden, MA: Wiley Blackwell.

Huber, Magnus and Kari Dako. 2004. Ghanaian English: morphology and syntax. In Kortmann and Schneider, eds.: 854-865.

Huber, Magnus. 2014. Stylistic variation in an emerging New English. Evidence from ICE-Ghana. Paper presented at the ICAME 35. University of Nottingham.

Huddleston, Rodney D. and Geoffrey K. Pullum, eds. *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Hundt, Marianne. 1998. *New Zealand English grammar, fact or fiction?: A corpus based study in morphosyntactic variation.* Amsterdam: John Benjamins.

Hundt, Marianne. 2012. Varieties of English: Australian/New Zealand English. In Bergs and Brinton, eds.: 1995–2012.

Hung, Tony T.N. 2012. Hong Kong English. In Low and Hashim, eds.: 113-133.

Hymes, Dell, ed. 1971. *Pidginization and creolization of languages.* Cambridge: Cambridge University Press.

Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar.* Cambridge, Mass.: MIT Press.

Jantos, Susanne. 2009. Agreement in educated Jamaican English: a corpus investigation of ICE-Jamaica. PhD thesis, University of Freiburg.

Jespersen, Otto. 1924. *The Philosophy of Grammar.* London: George Allen & Unwin.

Jespersen, Otto. 1933. *Linguistica: selected papers in English, French and German.* Copenhagen: Levin and Munksgaard.

Jespersen, Otto. 1949. *A modern English grammar on historical principles, Part VII. Syntax.* Copenhagen: Munksgaard.

Johansson, Stig. 1997. A corpus study of English existential clauses: register variation and discourse function. In Nevalainen and Kahlas-Tarkka, eds.: 303-318.

Johnson, Robert Keith. 1994. Language policy and planning in Hong Kong. *Annual Review of Applied Linguistics* 14. 177-199.

Kachru, Braj B. 1982. *The other tongue: English across cultures.* Urbana: University of Illinois Press.

Kachru, Braj B. 1988. The sacred cow of English. *English Today* 4: 3-8.

Kachru, Braj B., Yamuna Kachru and Cecil L. Nelson, eds. 2009. *The handbook of World Englishes*. Malden, MA: Wiley-Blackwell.
Kachru, Yamuna. 2006. *Hindi*. Amsterdam: John Benjamins.
Kallen, Jeffrey L. 1994. English in Ireland. In Burchfield, ed.: 148-196.
Kandiah, Thiru. 1998. Why New Englishes? In Foley, ed.: 1-40.
Kerswill, Paul and Ann Williams. 2000. Creating a new town koine: children and language change in Milton Keynes. *Language in Society* 29 (1): 65-115.
King, Robert D. 2006. First steps: Wales and Ireland. In Kachru, Kachru and Nelson, eds.: 30-40.
Kirkpatrick, Andy. 2012. Theoretical issues. In Low and Hashim, eds.: 13-31.
Kirkpatrick, Andy and Xu Zhichang. 2002. Chinese pragmatic norms and 'China English'. *World Englishes* 21 (2): 269-279.
Kortmann, Bernd and Edgar W. Schneider, eds. 2004. *A handbook of varieties of English: a multimedia reference tool. Volume 1: Phonology. Volume 2: Morphology and Syntax*. Berlin: De Gruyter. http://www.degruyter.com/viewbooktoc/product/36453 (accessed 19 April 2014).
Kortmann, Bernd und Benedikt Szmrecsanyi. 2004. Global synopsis: morphological and syntactic variation in English. In Kortmann and Schneider, eds.: 1142-1201.
Kortmann, Bernd and Kerstin Lunkenheimer, eds. 2012. *The Mouton world atlas of variation in English*. Berlin: De Gruyter.
Krishnaswamy, N. and Archana S. Burde. 1998. *The politics of Indians' English: linguistic colonialism and the expanding English empire*. Delhi: Oxford University Press.
Kuiper, Koenraad and Allan Bell. 2000. New Zealand and New Zealand English. In Bell and Kuiper, eds.: 11-22.
Kuno, Susumu. 1972. Functional sentence perspective: a case study from Japanese and English. *Linguistic Inquiry* 3 (3): 269-320.
Kuzar, Ron and Hadar Netz. 2010. Empirical support for the principle of the separation of reference and role. *Journal of Pragmatics* 42: 1460–1468.
Lalla, Barbara and Jean D'Costa. 1990. *Language in exile. Three hundred years of Jamaican Creole*. Tuscaloosa, Ala.: University of Alabama Press.
Lambrecht, Knud. 1994. *Information structure and sentence form : topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
Lambrecht, Knud. 2001a. Dislocation. In Haspelmath et al., eds.: 1050-1078.
Lambrecht, Knud. 2001b. A framework for the analysis of cleft constructions. *Linguistics* 39 (3): 463–516.
Lange, Claudia. 2007. Focus Marking in Indian English. *English World-Wide* 28 (1): 89-118.
Lange, Claudia. 2012. *The syntax of spoken Indian English*. Amsterdam: John Benjamins.
Lawton, David. 1980. Language attitude, discreteness, and code-shifting in Jamaican Creole. *English World-Wide* 1 (2): 211-226.
Lee, Nala Huiying, Ling Ai Ping and Hiroki Nomoto. 2009. Colloquial Singapore English *got*: functions and substratal influences. *World Englishes* 28 (3): 293-

318.
Leimgruber, Jakob R. E. 2013. *Singapore English : structure, variation and usage*. Cambridge: Cambridge University Press.
Li, Charles N., ed. 1976. *Subject and topic*. New York: Academic Press.
Li, Charles N. and Sandra A. Thompson. 1976. Subject and topic: a new typology of language. In Li, ed.: 457-489.
Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
Li, David C. S. 2000. 'Hong Kong English': new variety of English or interlanguage? *English Australia Journal* 18 (1): 50-59.
Li, Kening. 2008. Contrastive focus structure in Mandarin Chinese. In Marjorie K.M. Chan and Hana Kang, eds. *Proceedings of the 20th North American Conference on Chinese Linguistics* (NACCL-20). Columbus, Ohio: The Ohio State University. 759-774.
Lick, Ho Chee and Lubna Alsagoff. 1998. English as the common language in multicultural Singapore. In Foley, ed.: 201-217.
Lim, Lisa, ed. 2004. *Singapore English. A grammatical description*. Amsterdam: John Benjamins.
Lim, Lisa and Joseph A. Foley. 2004. English in Singapore and Singapore English. In Lim, ed.: 1-18.
Lodge, Ken R. 1979. A three-dimensional analysis of non-standard English. *Journal of Pragmatics* 3: 169-195.
Los, Bettelou and Erwin Komen. 2012. Clefts as resolution strategies after the loss of a multifunctional first position. In Nevalainen and Traugott, eds.: 884-898.
Low, Ee Ling and Azirah Hashim, eds. 2012. *English in Southeast Asia: features, policy and language in use*. Amsterdam: John Benjamins.
Mair, Christian. 1992. Problems in the compilation of a corpus of standard Caribbean English: a pilot study. In Leitner, ed. *New directions in English language corpora: methodology, results, software developments*. Berlin: De Gruyter. 75-96.
Mair, Christian. 1997. Parallel Corpora: a Real-Time Approach to the Study of Language Change in Progress. In Magnus Ljung, ed. *Corpus-Based Studies in English*. Amsterdam: Rodopi. 195-209.
Mair, Christian 2002a. Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora. *English Language and Linguistics* 6 (1): 105-131.
Mair, Christian 2002b. Creolisms in an emerging standard: written English in Jamaica. *English World-Wide* 23 (1): 31–58.
Mair, Christian. 2013. Using 'small' corpora to document ongoing grammatical change. In Manfred Krug and Julia Schlüter, eds. *Research methods in language variation and change*. Cambridge: Cambridge University Press. 181-194.
Mair, Christian and Claudia Winkle. 2012. Change from *to*-infinitive to bare infinitive in specificational cleft sentences: data from World Englishes. In Marianne Hundt and Ulrike Gut, eds. *Mapping Unity and Diversity World-Wide. Corpus-Based Studies of New Englishes*. Amsterdam: John Benjamins. 243-262.
Martin, Isabel Pefianco. 2014. Beyond nativization? Philippine English in Schneider's Dynamic Model. In Buschfeld et al., eds.: 70-85.

Martinez Insua, Ana E. and Ignacio M. Palacios Martinez. 2003. A corpus-based approach to non-concord in present day English existential *there*-constructions. *English Studies* 84 (3): 262–83.

Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.

Matras, Yaron and Jeanette Sakel. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31 (4): 829–865.

Matthews, Stephen and Virginia Yip. 1994. *Cantonese: a comprehensive grammar*. London: Routledge.

McArthur, Tom. 1987. The English languages? *English Today* 3 (3): 9-13.

McCarthy, Michael and Ronald Carter. 1997. Grammar, tails, and affect: Constructing expressive choices in discourse. *Text* 17 (3): 405-429.

McCartney, Donal. 1987. *The dawning of democracy : Ireland 1800-1870*. Dublin: Helicon.

McCloskey, James. 2012. Irish existentials in context. To appear in *Syntax*.

McNally, Louise. 1997. *A semantics for the English existential construction*. New York: Garland.

McWhorter, John. 2006. Creole transplantation: a source of solutions to resistant anomalies. In J. Clancy Clements, Thomas A. Klingler, Deborah Piston-Hatlen and Kevin J. Rottet, eds. *History, society and variation. In honour of Albert Valdman*. Amsterdam: John Benjamins. 103-133.

Meechan, Marjory and Michele Foley. 1994. On resolving disagreement: linguistic theory and variation - *there's bridges*. *Language Variation and Change* 6: 63–85.

Mehrotra, Raja Ram. 1998. *Indian English: texts and interpretations*. Amsterdam: John Benjamins.

Melchers, Gunnel. 1983. It's a sweet thing, is tea-cake — a study of tag statements. In Sven Jacobson, ed. *Papers from the Second Scandinavian Symposium on Syntactic Variation*. Stockholm: Acta Universitatis Stockholmiensis. 57–66.

Melchers, Gunnel and Philip Shaw. 2003. *World Englishes*. London: Arnold.

Mesthrie, Rajend. 1992. *English in language shift. The history, structure, and sociolinguistics of South African Indian English*. Cambridge: Cambridge University Press.

Mesthrie, Rajend. 2008. English circling the globe. *English Today* 24 (1): 28-32.

Mesthrie, Rajend and Rakesh M. Bhatt. 2008. *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Meyerhoff, Miriam. 1994. Sounds pretty ethnic, eh? A pragmatic particle in New Zealand English. *Language in Society* 23 (3): 367-388.

Michaelis, Laura A. and Knud Lambrecht. 1996. Toward a construction-based theory of language function: the case of nominal extraposition. *Language* 72 (2): 215-247.

Miller, Jim. 1996. Clefts, particles and word order in languages in Europe. *Language Sciences* 18: 111-125.

Miller, Jim. 2006. Focus. In Brown, ed.: 511-518.

Miller, Jim and Regina Weinert. 1998. *Spontaneous spoken language. Syntax and discourse*. Oxford: Clarendon Press.

Milsark, Gary L. 1979. *Existential Sentences in English.* New York: Garland.

Ministry of Education Singapore. 2009. *English language syllabus 2010: primary & secondary.* http://www.moe.gov.sg/education/syllabuses/ (accessed 5 August 2014).

Mufwene, Salikoko S. 2001. *The ecology of language evolution.* Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2008. *Language evolution: contact, competition and change.* London: Continuum.

Mukherjee, Joybrato. 2007. Steady states in the evolution of New Englishes: present-day Indian English as an Equilibrium. *Journal of English Linguistics* 35: 157-187.

Nagaya, Naonori. 2007. Information structure and constituent order in Tagalog. *Language and Linguistics* 8 (1): 343-372.

Netz, Hadar and Ron Kuzar. 2007. Three marked theme constructions in spoken English. *Journal of Pragmatics* 39 (2): 305–335.

Netz, Hadar, Ron Kuzar and Zohar Eviatar. 2011. A recipient-based study of the discourse functions of marked topic constructions. *Language Sciences* 33 (1): 154–166.

Nevalainen, Terttu and Leena Kahlas-Tarkka, eds. 1997. *To explain the present: studies in the changing English language in honour of Matti Rissanen.* Helsinki, Finland: Société néophilologique.

Nevalainen, Terttu and Elizabeth Cross Traugott, eds. 2012. *The Oxford handbook of the history of English.* Oxford: Oxford University Press.

Newbrook, Mark. 1988. Relative clauses, relative pronouns and Hong Kong English. University of Hong Kong Papers on Linguistics and Language Teaching 11: 25-41.

Newbrook, Mark. 1998. Which way? That was? Variation and ongoing changes in the English relative clause. *World Englishes* 17 (1): 43-59.

Oberlander, Jon and Judy Delin. 1996. The function and interpretation of reverse *wh*-clefts in spoken discourse. *Language and Speech* 39 (2/3): 185-227.

Ortega, Lourdes. 2009. *Second language acquisition.* London: Hodder Education.

Parviainen, Hanna. 2012. Focus particles in Indian English and other varieties. *World Englishes* 31 (2): 226-247.

Patrick, Peter L. 1999. *Urban Jamaican Creole: variation in the mesolect.* Amsterdam: John Benjamins.

Patrick, Peter L. 2004. Jamaican Creole: morphology and syntax. In Kortmann and Schneider, eds.: 407-438.

Patrick, Peter L. 2007. Jamaican Patwa (Creole English). In John Holm and Peter L. Patrick, eds. *Comparative creole syntax: parallel outlines of 18 creole grammars.* London: Battlebridge. 127-152.

Patrick, Peter L. 2012. Jamaican Creole. In Kortmann and Lunkenheimer, eds.: 222-236.

Patten, Amanda. 2012a. *The English 'it'-cleft. A constructional account and a diachronic investigation.* Berlin: De Gruyter.

Patten, Amanda. 2012b. The historical developement of the *it*-cleft: a comparison of two different approaches. *Studies in Language* 36 (3): 548-575.

Paul, M., Gary F. Simons and Charles D. Fennig, eds. 2014. *Ethnologue: Languages of the World*, Seventeenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Paul, Waltraud and John Whitman. 2008. *Shi…de* focus clefts in Mandarin Chinese. *The Linguistic Review* 25 (3/4): 413-451.

Pfenninger, Simone E. 2009. *Grammaticalization paths of English and High German existential constructions: a corpus-based study*. Bern: Peter Lang.

Phillips, Jock. 2013. History of immigration. In *Te Ara - the Encyclopedia of New Zealand*. http://www.teara.govt.nz/en/history-of-immigration (accessed 29 October 2014).

Pietsch, Lukas. 2008. Prepositional aspect constructions in Hiberno-English. In Siemund and Kintana, eds.: 213-236.

Platt, John Talbot, Heidi Weber and Mian Lian Ho. 1983. *Singapore and Malaysia*. Amsterdam: John Benjamins.

Platt, John Talbot, Heidi Weber and Mian Lian Ho. 1984. *The New Englishes*. London: Routledge and Kegan Paul.

Poutsma, Hendrik. 1928. *Grammar of Late Modern English*. Groningen: P. Noordhoff.

Preusler, Walther. 1938. Keltischer Einfluss im Englischen. Indogermanische Forschungen 56 (1): 178-191.

Preusler, Walther. 1942. Keltischer Einfluss im Englischen. *Anglia – Zeitschrift für englische Philologie* 66: 121-128.

Primus, Beatrice. 1993. Word order and information structure: a performance-based account of topic positions and focus positions. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann, eds. *Syntax. An international handbook of contemporary research*. Berlin: De Gruyter. 880-896.

Prince, Ellen F. 1978. Comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54 (4): 883-906.

Prince, Ellen F. 1981. Toward a taxonomy of given/new information. In Peter Cole, ed. *Radical Pragmatics*. New York: Academic Press. 223-254.

Prince, Ellen F. 1985. Fancy syntax and 'shared knowledge'. *Journal of Pragmatics* 9: 65-81.

Prince, Ellen F. 1992. The ZPG Letter: subjects, definiteness, and information-status. In S. A. Thompson and W. C. Mann, eds. *Discourse Description: Diverse Analyses of a Fundraising Text*. Amsterdam: John Benjamins. 295–325.

Prince, Ellen F. 1998. On the limits of syntax, with reference to left-dislocation and topicalization. In Peter Culicover and Louise McNally, eds. *Syntax and semantics: the limits of syntax* 29. San Diego: Academic Press. 281-302.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Quirk, Randolph. 1990. Language varieties and standard language. *English Today* 6 (1): 3-10.

Rafael, Teresita C. 1978. Topic in Tagalog revisited. *Studies in Philippine Linguistics* 2 (1): 36-48.

Reeve, Matthew. 2012. *Clefts and their relatives*. Amsterdam: John Benjamins.

Reinhart, Tanya. 1981. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27 (1): 53–94.

Rickford, John R. 1986. Social contact and linguistic diffusion: Hiberno-English and New World Black English. *Language* 62 (2): 245-289.

Rickford, John R., Thomas A. Wasow, Norma Mendoza-Denton and Juli Espinoza. 1995. Syntactic variation and change in progress: loss of the verbal coda in topic-restricting *as far as* constructions. *Language* 71 (1): 102-131.

Ritchie, William C. 1986. Second language acquisition research and the study of non-native varieties of English: some issues in common. *World Englishes* 5 (1): 15-30.

Rochemont, Michael Shaun and Peter W. Culicover. 1990. *English focus constructions and the theory of grammar.* Cambridge: Cambridge University Press.

Rohdenburg, Günter. 1998. Clarifying Structural Relationships in Cases of Increased Complexity in English. In Rainer Schulze, ed. *Making meaningful choices in English: On dimensions, perspectives, methodology and evidence.* Tübingen: Narr. 177-205.

Rohdenburg, Günter. 2000. The complexity principle as a factor determining grammatical variation and change in English. In Ingo Plag and Klaus Peter Schneider, eds. *Language use, language acquisition and language history: (mostly) empirical studies in honour of Rüdiger Zimmermann.* Trier: WVT. 25-44.

Rohdenburg, Günter and Julia Schlüter. 2009. *One language, two grammars? Differences between British and American English.* Cambridge: Cambridge University Press.

Rosenfelder, Ingrid. 2009. Sociophonetic variation in educated Jamaican English: an analysis of the spoken component of ICE-Jamaica. PhD thesis, University of Freiburg.

Rudanko, Juhani. 1998. *Change and continuity in the English language : studies on complementation over the past three hundred years.* Lanham, MD: University Press of America.

Rudanko, Juhani. 2000. *Corpora and complementation: tracing sentential complementation patterns of nouns, adjectives and verbs over the last three centuries.* Lanham, MD: University Press of America.

Rühlemann, Christoph. 2006. Coming to terms with conversational grammar: 'dislocation' and 'dysfluency'. *International Journal of Corpus Linguistics* 11 (4): 385–409.

Rupp, Laura. 2005. Constraints on nonstandard *-s* in expletive *there* sentences: A generative-variationist perspective. *English Language and Linguistics* 9 (2): 255–88.

Sabbagh, Joseph. 2009. Existential sentences in Tagalog. *Natural Language and Linguistic Theory* 27 (4): 675-719.

Sailaja, Pingali. 2009. *Indian English.* Edinburgh: Edinburgh University Press.

Sand, Andrea. 1999. *Linguistic variation in Jamaica. A corpus-based study of radio and newspaper usage.* Tübingen: G. Narr.

Schachter, Paul. 1973. Focus and relativization. *Language* 49 (1): 19-46.

Schachter, Paul and Fe T. Otanes. 1972. *Tagalog reference grammar.* Berkeley/LA/London: University of California Press.

Schachter, Jacquelyn and Marianne Celce-Murcia. 1977. Some reservations concerning Error Analysis. *TESOL Quarterly* 11 (4): 441-451.

Schilling-Estes, Natalie and Walt Wolfram. 1994. Convergent explanation and alternative regularization patterns: *were/weren't* leveling in a vernacular English variety. *Language Variation and Change* 6: 273–302.
Schmied, Josef. 1996. Second-language corpora. In Greenbaum, ed. 182-196.
Schneider, Edgar W. 2003. The dynamics of New Englishes: from identity construction to dialect birth. *Language* 79 (2): 233-281.
Schneider, Edgar W. 2007. *Postcolonial English. Varieties around the world.* Cambridge: Cambridge University Press.
Schneider, Edgar W. 2014. New reflections on the evolutionary dynamics of world Englishes. *World Englishes* 33 (1): 9-32.
Schreier, Daniel. 2003. Convergence and language shift in New Zealand: consonant cluster reduction in 19th century Maori English. *Journal of Sociolinguistics* 7 (3): 378–391.
Schütze, Carson T. 1999. English expletive constructions are not infected. *Linguistic Inquiry* 30 (3): 467–84.
Sedlatschek, Andreas. 2009. *Contemporary Indian English: variation and change.* Amsterdam: John Benjamins.
Senior, Olive. 2003. Encyclopedia of Jamaican Heritage. St. Andrew, Jamaica: Twin Guinep Publishers.
Seoane, Elena. 2006. Information structure and word order change: the passive as an information-rearranging strategy in the history of English. In Van Kemenade and Los, eds.: 360-391.
Setter, Jane, Cathy S. P. Wong and Brian H. S. Chan. 2010. *Dialects of English: Hong Kong English.* Edinburgh: Edinburgh University Press.
Sharma, Devyani. 2012a. Shared features in New Englishes. In Hickey, ed.: 211-232.
Sharma, Devyani. 2012b. Indian English. In Kortmann and Lunkenheimer, eds.: 523-530.
Shields, Kathryn. 1989. Standard English in Jamaica: a case of competing models. *English World-Wide* 10 (1): 41-53.
Siemund, Peter and Noemi Kintana, eds. 2008. *Language contact and contact languages.* Amsterdam: John Benjamins.
Siemund, Peter and Kalynda Beal. 2011. Clefting in contact through space and time: an analysis of *it*-clefts in Irish English, Indian English and British English. In Raymond Hickey, ed. *Researching the languages of Ireland.* Uppsala: Uppsala University Press.
Sinha, R. Mahesh K. and Anil Thakur. 2005. Translation divergence in English-Hindi MT. EAMT 2005 Conference Proceedings.
Sornicola, Rosanna. 1988. *It*-clefts and *wh*-clefts: two awkward sentence types. *Journal of Linguistics* 24 (2): 343-379.
Speyer, Augustin. 2012. Stress and word order changes in the left periphery in Old English and Middle English. In Nevalainen and Traugott, eds.: 873-883.
Statistics Canada. 2012. 2011 Census of Population: immigrant languages in Canada. http://www12.statcan.gc.ca (accessed 4 November 2014).
Statistics New Zealand. 2013. Results of a census of the colony of New Zealand taken for the night of the 28th March 1886. http://www.stats.govt.nz (accessed 29 October 2014).

Statistics New Zealand. 2014. 2013 Census QuickStats about culture and identity. Published 15 April 2014. http://www.stats.govt.nz (accessed 29 October 2014).

Stenson, Nancy. 1981. *Studies in Irish syntax*. Tübingen: Narr.

Tagliamonte, Sali. 2002. Variation and change in the British relative marker system. In Patricia Poussa, ed. *Dialect contact on the North Sea Littoral*. Munich: Lincom Europa. 147-165.

Tagliamonte, Sali A. 2006. Historical change in synchronic perspective: the legacy of British dialects. In Van Kemenade and Los, eds.: 477-506.

Tang, Sze-Wing and Siu-Pong Cheng. 2014. Aspects of Cantonese grammar. In Huang, Li and Simpson, eds.: 601-628.

Tay, Mary W. J. 1991. Southeast Asia and Hongkong. In Jenny Cheshire, ed. *English around the world: sociolinguistic perspectives*. Cambridge: Cambridge University Press. 319-332.

The Government of the Hong Kong Special Administrative Region. 1990. *The basic law of the Hong Kong Special Administrative Region of the People's Republic of China*. http://www.basiclaw.gov.hk/en/basiclawtext/images/basiclaw_full_text_en.pdf (accessed 14 September 2014).

The Government of the Hong Kong Special Administrative Region, Census and Statistics Department. 2001. *2001 population census: summary results*. http://www.statistics.gov.hk/pub/B11200182001XXXXB0100.pdf (13 September 2014).

The Government of the Hong Kong Special Administrative Region, Census and Statistics Department. 2012. *2011 population census: summary results*. http://www.census2011.gov.hk/pdf/summary-results.pdf (accessed 13 September 2014).

Thomason, Sarah Grey. 2001. *Language contact : an introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey. 2010. Contact explanations in Linguistics. In Hickey, ed.: 31-47.

Thomason, Sarah Grey and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Thompson, Roger M. 2003. *Filipino English and Taglish: language switching from multiple perspectives*. Amsterdam: John Benjamins.

Timmis, Ivor. 2010. 'Tails' of Linguistic Survival. *Applied Linguistics* 31 (3): 325–345.

Tottie, Gunnel. 1997. Relatively speaking. Relative marker usage in the British National Corpus. In Nevalainen and Kahlas-Tarkka, eds.: 465-481.

Tottie, Gunnel and Michel Rey. 1997. Relativization strategies in earlier African American Vernacular English. *Language Variation and Change* 9 (2): 219-247.

Vallduví, Enric. 1994. Detachment in Catalan and information packaging. *Journal of Pragmatics* 22: 573-601.

Vallduví, Enric and Elisabet Engdahl. 1996. The linguistic realisation of information packaging. *Linguistics* 34: 459-519.

Van Kemenade, Ans and Bettelou Los, eds. 2006. *The handbook of the history of English*. Malden, MA: Blackwell.

Visser, Fredericus Th. 1963. *An Historical Syntax of the English Language*. Leiden: Brill.

Walker, James A. 2007. "There's bears back there". Plural existentials and vernacular universals in (Quebec) English. *English World-Wide* 28 (2): 147-166.

Ward, Gregory L. and Ellen F. Prince. 1991. On the topicalization of indefinite NPs. *Journal of Pragmatics* 16 (2): 167-177.
Ward, Gregory and Betty Birner. 1995. Definiteness and the English existential. *Language* 71 (4): 722-742.
Ward, Gregory, Betty Birner and Rodney D. Huddleston. 2002. Information packaging. In Huddleston and Pullum, eds.: 1363-1447.
Ward, Gregory and Betty Birner. 2004. Information structure and non-canonical syntax. In L.R. Horn and G. Ward, eds. *The Handbook of Pragmatics*. Oxford and Malden, MA: Blackwell. 153–174.
Warren, Tessa and Edward Gibson. 2005. Effects of NP type in reading cleft sentences in English. *Language and Cognitive Processes* 20 (6): 751-767.
Weinert, Regina and Jim Miller. 1996. Cleft constructions in spoken language. *Journal of Pragmatics* 25: 173-206.
Weinreich, Uriel. 1968. *Languages in contact : findings and problems*. The Hague: Mouton.
Williams, Jessica. 1987. Non-native varieties of English: a special case of language acquisition. *English World-Wide* 8 (2): 161-199.
Wilson, John. 2014. European discovery of New Zealand. In *Te Ara - the Encyclopedia of New Zealand*. http://www.teara.govt.nz/en/european-discovery-of-new-zealand (accessed 29 October 2014).
Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Oxford and Malden, MA: Blackwell.
Winkle, Claudia. 2011. Pseudo-cleft constructions in varieties of English: a study based on ten ICE corpora. Unpublished qualification thesis, University of Freiburg.
Winkler, Susanne. 1997. *Focus and secondary predication*. Berlin: De Gruyter.

Corpora

Davies, Mark. 2008-. *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.
Davies, Mark. 2013. *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries*. Available online at http://corpus.byu.edu/glowbe/.
International Corpus of English. http://ice-corpora.net/ice/. Last updated: 15 September 2015.

Software

The R Foundation for Statistical Computing. 2014. R version 3.1.1. Available online at http://www.r-project.org/.

Zusammenfassung in deutscher Sprache

Die vorliegende Studie untersucht nicht-kanonische Satzstrukturen in verschiedenen Varietäten des Englischen. Diese Strukturen werden verwendet, um von der eher fixen Wortstellung in der englischen Sprache abzuweichen. In kanonischen Sätzen steht in der Regel das Subjekt am Satzanfang, gefolgt vom Prädikat. Manchmal wollen Sprecher allerdings von dieser kanonischen Satzstruktur abweichen, um dem Gesprächspartner bestimmte Informationen auf eine bestimmte Art vermitteln zu können. So verwenden sie nicht-kanonische Strukturen zum Beispiel, um neue Informationen - Gegenstände, Personen oder Ereignisse - in das Gespräch einzuführen, um früher bereits erwähnte Informationen später im Gespräch wieder aufzugreifen oder um bestimmte Informationen als besonders wichtig und relevant hervorzuheben. Derartige Strukturen sind Gegenstand der vorliegenden Studie. Sie untersucht Linksversetzungen (*left dislocation*), Rechtsversetzungen (*right dislocation*), Vorfeldbesetzung (*fronting*), Existenzsätze mit *there + be* (*existential* there-*constructions*) und verschiedene Formen von Spaltsätzen (*cleft constructions*). Beispiele dieser Strukturen sind im Folgenden aufgeführt:

a. Tom, he loves apples. - 'Tom, der liebt Äpfel.'
b. He loves apples, Tom. - 'Er liebt Äpfel, Tom.'
c. Apples Tom loves. - 'Äpfel liebt Tom.'
d. There are apples on the table. - 'Da sind Äpfel auf dem Tisch.'
e. It's Tom who loves apples. - 'Es ist Tom, der Äpfel liebt.'

Frühere Studien haben gezeigt, dass es quantitative und qualitative Unterschiede gibt im Gebrauch dieser Strukturen und dass Sprecher verschiedener englischer Dialekte/Varietäten verschiedene Strukturen bevorzugen. Diese Studien haben wertvolle Einblicke geliefert in die Form und den Gebrauch der Strukturen, aber ein Vergleich der Ergebnisse ist schwierig, da die Studien zum Teil auf unterschiedlichen Definitionen der Strukturen basieren und sehr verschiedenes Datenmaterial verwenden. Hier setzt die vorliegende Studie an. Ziel dieser Arbeit ist es, einen umfassenden Überblick über die oben genannten nicht-kanonischen Strukturen zu geben und Antworten auf die folgenden Fragen zu finden: Unterscheiden sich Sprecher verschiedener englischer Varietäten in der Art, wie sie die Informationen in einem Satz strukturieren? Das heißt, haben sie Präferenzen im Gebrauch von *left dislocation*, *right dislocation*, *fronting*, *existential* there-*constructions* und *clefts*? Wenn es Unterschiede gibt,

sind diese eher von quantitativer oder qualitativer Art? Erfüllen die Konstruktionen dieselben Funktionen in den verschiedenen englischen Varietäten? Was motiviert die Sprecher, nicht-kanonische Strukturen zu verwenden? Um Antworten auf diese Fragen zu finden, werden die Form und der Gebrauch der oben genannten Strukturen in verschiedenen Varietäten des Englischen untersucht und verglichen. Als Datengrundlage dienen neun Komponenten des *International Corpus of English* (ICE). Genauer gesagt, werden die oben genannten Strukturen in informellen gesprochenen Texten aus England, Irland, Neuseeland, Kanada, Singapur, den Philippinen, Jamaika, Indien und Hongkong untersucht, wobei die Texte für jede Varietät ungefähr 200 000 Wörter umfassen. Einzelne Analysen werden zudem zusätzlich im *Corpus of Global Web-based English* (GloWbE) und/oder im *Corpus of Contemporary American English* (COCA) durchgeführt, da die geringen Datenmengen der ICE Korpora häufig keine zuverlässigen Schlussfolgerungen zulassen. GloWbE umfasst ungefähr 1,9 Milliarden Wörter, die von Webseiten aus 20 verschiedenen Ländern stammen. Der Großteil der Texte besteht aus informellen Blogs, zudem wurden formalere Zeitungstexte oder Webseiten von Firmen in das Korpus aufgenommen. COCA umfasst ungefähr 450 Millionen Wörter aus verschiedenen Genres (z.B. akademische Texte, gesprochene Texte, Zeitungstexte) (siehe Kapitel 1).

Die historische Entwicklung und aktuelle Situation der englischen Sprache in den Ländern, deren Varietäten in der vorliegenden Studie untersucht werden, werden in Kapitel 2 beschrieben. Ein derartiges Hintergrundwissen kann hilfreich sein, um sprachliches Verhalten besser einordnen und verstehen zu können.

Grundlegende theoretische Begriffe zum Thema Informationsstruktur werden in Kapitel 3 diskutiert. Außerdem werden die Formen und die Funktionen von *left dislocation*, *right dislocation*, *fronting*, *existential* there-*constructions* und *clefts* beschrieben, sowie frühere Studien zusammengefasst, die diese Strukturen untersuchen. Die Struktur normaler Sätze in einigen Muttersprachen (z.B. Irisch, Hindi, Mandarin) und äquivalente Formen zu den englischen nicht-kanonischen Strukturen werden zudem in diesem Kapitel kurz erläutert, da Transferphänomene so besser erkannt und erklärt werden können.

Die Ergebnisse der vorliegenden Studie zeigen, dass Sprecher verschiedener englischer Varietäten in der Tat unterschiedliche nicht-kanonische Strukturen bevorzugen (Kapitel 4 und 5). Besonders auffällig ist der hochfrequente Gebrauch von *left dislocation* und *fronting* im indischen Englisch im Vergleich zu den acht anderen englischen Varietäten. Da keine strukturellen oder funktionalen Besonderheiten festgestellt werden können, kommt die Studie zu dem Ergebnis, dass diese Strukturen im indischen Englisch schwächer markierte grammatische Erscheinungen sind, eventuell zum Repertoire der kanonischen Satzstrukturen gehören und deshalb häufiger verwendet werden als in anderen Varietäten. Es erscheint plausibel, dass Einfluss von indischen Sprachen zu dieser Entwicklung beigetragen hat, da diese Sprachen in kanonischen Sätzen das Objekt vor das Verb stellen und im Bereich der Wortstellung flexibler sind als das Englische.

Weitere interessante Ergebnisse können für irisches Englisch festgestellt werden. Sprecher dieser englischen Varietät verwenden *left dislocation*, *right dislocation* und *it-clefts* häufiger als die Muttersprachler aus England, Neuseeland und Kanada, und sie zeigen mehr Variation in ihrer Form und ihrem Gebrauch. Das besondere Ver-

halten lässt sich in vielen Fällen auf Einfluss des Irischen auf das Englische zurückführen und auf die Tatsache, dass Irland eine englische Kolonie war und die englische Sprache deshalb als Zweitsprache oft von Erwachsenen und innerhalb eines sehr kurzen Zeitraumes erworben wurde.

Einige der Sprecher, die Englisch als Zweitsprache erworben haben, haben eine topik-prominente Muttersprache. Das bedeutet, dass Sätze nicht auf dem grundlegenden Prinzip von Subjekt und Prädikat aufbauen wie im Englischen, sondern dass das Konzept des Topiks eine bedeutendere Rolle spielt. Chinesische Sprachen und viele der Sprachen, die auf den Philippinen gesprochen werden, gehören zur Familie der topik-prominenten Sprachen. Der Einfluss dieser Sprachen auf das Englische lässt sich in verschiedenen Bereichen erkennen. Besonders interessant ist die Entwicklung einer speziellen Form von *left dislocation*, in der die Präposition *for* als Topikmarker fungiert (***For** me, I don't like apples*). Diese Struktur wird vor allem von englischen Sprechern aus Singapur und Hongkong verwendet, gefolgt von Sprechern aus den Philippinen. Vermutlich versuchen die Sprecher, die englischen Sätze nach den Prinzipien ihrer Muttersprache zu strukturieren. Außerdem sind den Sprechern die englischen Strukturen mit *as for* und *for NP* bekannt, die zur Einführung einer Person oder eines Gegenstandes als Topik dienen oder die verwendet werden, um eine Meinung kundzutun, oft im Vergleich zur Meinung einer anderen Person. Diese drei Faktoren - topik-prominente Muttersprache, Kenntnis der englischen Strukturen mit *as for* und *for NP* - scheinen also zur Grammatikalisierung von *for* in einen Topikmarker beizutragen.

Interessante Unterschiede können außerdem im Vergleich von Singapur Englisch und Hong-kong Englisch festgestellt werden. Wie bereits erwähnt, haben die beiden englischen Varietäten gemeinsam, dass sie in Kontakt mit chinesischen Sprachen kommen. Allerdings haben die Sprecher unterschiedlich gute Kenntnisse in der englischen Sprache. Während man im Hong-kong Englisch einige Merkmale findet, die die Sprecher eindeutig als Lerner der Sprache auszeichnen, findet man im Singapur Englisch Strukturen, die zeigen, dass die englische Sprache hier schon eigene, lokale Normen entwickelt hat und sich somit vom britischen Muster abhebt. Ein Beispiel ist das Verb *got*, das verwendet wird, um Existenz auszudrücken (z.B. *Inside cake got fruit ~There's fruit in the cake*). Die geringere Häufigkeit an Existenzsätzen mit *there + be* in Singapur Englisch lässt sich vermutlich auf das Vorkommen dieser Struktur zurückführen. Derartige Beispiele zeigen, dass die englische Sprache in Singapur institutionalisiert ist und viele Sprecher ein hohes Maß an Fähigkeit aufweisen.

Aus methodischer Sicht hat es sich als hilfreich erwiesen, in manchen Fällen die Analyse der ICE Daten durch Daten von GloWbE und/oder COCA zu ergänzen. So konnten einzelne Vermutungen, die nur auf geringen Datenmengen basierten, bestärkt werden. Einige Strukturen erwiesen sich allerdings als sehr selten in GloWbE. Das zeigt, dass manche Strukturen, die Gegenstand der vorliegenden Studie sind, eindeutig Phänomene von informeller gesprochener Sprache sind.

This book contributes to the study of information structure in varieties of English around the world. Situated at the interface of dialectology, syntax and pragmatics, it examines structural devices such as left and right dislocation, fronting, existential *there*-constructions and cleft sentences, which help speakers change the rather fixed word order of present-day English and organize sentence information in ways which are better suited to the discourse context.

Approaching information structure from a cross-varietal perspective, this study compares speech data from nine varieties of English, culled from the International Corpus of English and complemented by data from the Corpus of Global Web-based English and the Corpus of Contemporary American English. In its breadth and systematicity of coverage, this is the most comprehensive study to date of the use of information-packaging constructions across varieties of English. Through this comprehensive approach diverging patterns of use and frequency distributions can be identified. Factors impacting on the frequency of use involve substrate influence in language contact situations, universal developmental processes in language acquisition and specific features of the socio-cultural setting.

Claudia Winkle studied English and Mathematics at the University of Freiburg (Germany) and University College Dublin (Ireland), obtaining a teacher's diploma in both subjects in 2011. From 2011 to 2014 she was a member of the Freiburg team of the German CLARIN-D ("Common Language Resources and Technology Infrastructure") project and from 2013 to 2014 an associate member of the Freiburg-based doctoral research training group on "Frequency Effects in Language" (DFG-GRK 1624). She received her doctoral degree from the University of Freiburg in 2015. This book is based on her dissertation.

ISBN 978-3-928969-62-8

UNI FREIBURG