Varieties of English Around the World G45

The Syntax of Spoken Indian English

Claudia Lange

John Benjamins Publishing Company

The Syntax of Spoken Indian English

Varieties of English Around the World (VEAW)

A companion monograph series devoted to sociolinguistic research, surveys and annotated text collections. The VEAW series is divided into two parts: a text series contains carefully selected specimens of Englishes documenting the coexistence of regional, social, stylistic and diachronic varieties in a particular region; and a general series which contains outstanding studies in the field, collections of papers devoted to one region or written by one scholar, bibliographies and other reference works.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/veaw

Volume G45

The Syntax of Spoken Indian English
by Claudia Lange

The Syntax of Spoken Indian English

Claudia Lange
University of Giessen

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Lange, Claudia.
The syntax of spoken Indian English / by Claudia Lange.
p. cm. (Varieties of English Around the World, ISSN 0172-7362 ; v. G45)
Includes bibliographical references and index.
1. English language--Spoken English--India. 2. English language--India--Syntax.
3. Languages in contact--India. 4. Interference (Linguistics) 5. Sociolinguistics--India. I. Title.
PE3502.I6L34 2012
427'.954--dc23 2012031474
ISBN 978 90 272 4905 0 (Hb ; alk. paper)
ISBN 978 90 272 7309 3 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

List of figures and tables

List of abbreviations

AA	Austro-Asiatic
AEV	Asian English variety
AmE	American English
BrE	British English
CIEFL	Central Institute for the Study of English and Foreign Languages, Hyderabad
CIIL	Central Institute for the Study of Indian Languages, Mysore
Dr	Dravidian
EFL	English as a foreign language
ENL	English as a native language
ESL	English as a second language
IA	Indo-Aryan
ICE	International Corpus of English
IE	Indo-European
IndE	Indian English
InSAfE	South African Indian English
IVE	Indigenized variety of English
L1	First language
L2	Second language
LD	Left dislocation
NIA	New Indo-Aryan
NNIVE	Non-native institutionalized variety of English
RD	Right dislocation
SAIE	South African Indian English
SLA	Second language acquisition
TB	Tibeto-Burman

Acknowledgments

My first ever visit to India in 2004 took me to Pune, where I spent three months as a guest lecturer in the English department of the University of Pune. It did not take me long to discover a noticeable gap between the constitutional status of English in India as a second official language and the realities out in the streets. Since then, my understanding of English in India has certainly improved, and both the dynamism of contemporary Indian society as well as the aspirations of millions of its people are changing the way English in India is used and perceived.

Many friends have made frequent visits to India a wonderful experience – and have made me want to come back many times: Karuna Kushare, Kasturi Dadhe, Koyna Sinha-Poduval, Lina Mukhopadhyay, Madhumeeta Sinha, Noel Arokianathan, Paritosh Sarda, Pradnya Dasarwar, Sachin Labade, Satish Poduval, Smriti Ranjan all cheerfully integrated me into their lives when I happened to be around, enlightening me about many aspects of Indian life, culture, politics – not to forget food and cricket!

The English and Foreign Languages University (EFL-U) Hyderabad has always been a most helpful host and starting point for further forays into the topic. I would like to thank its dedicated staff on all levels, from the guards in front of the guest houses to the lecturers in various departments, for providing such a stimulating environment.

My old home university, the Technical University Dresden, has not only benevolently tolerated, but also generously supported my rather far-flung research interest. Special thanks go first and foremost to Ursula Schaefer, who found more ways of encouraging me over many years than anybody can reasonably expect.

I am heavily indebted to Edgar Schneider and Steffi Hackert for their detailed critical comments on the draft version. Their suggestions were instrumental in writing a better book.

Many thanks go to Christof Müller (Dresden), who wrote a Python script that enabled me to carry out the sociolinguistic analyses in Chapter 4. I thank Sandra Weiser and Tobias Bernaisch (Giessen) for improving the final layout. Any remaining quirks, oddities and misconceptions are of course my own.

Claudia Lange Berlin, May 2012

CHAPTER 1

Introduction

1.1 English in India or Indian English?

The very notion of "Indian English" (IndE) has been fiercely contested both by linguists who happen to be interested in New Englishes and by more or less everyone concerned with the teaching of English in India. Braj B. Kachru claims that English is "an Asian Language", "a part of our local pluralistic linguistic heritage" (1998: 91). Many Indians speak English at the level of functional nativeness, and so there is definitely a case for considering Indian English as a variety in its own right as far as Kachru is concerned.[1] Others are more doubtful about the existence of a supraregional or even national variety to be labelled "Indian English":

> From India, Kachru (1986) has given credence to the idea that there exists an 'institutionalized second-language variety of English' called Indian English. This idea is confusing because it begs the question: which institute grants the certificate to 'institutionalize' it? This has never been fully substantiated because Indian English, if such a linguistic species actually exists, will be in at least 18 different varieties that are the regional standards of India. The term 'Indian English' misleads readers into thinking that this is one monolithic whole. (Vaish 2005: 190)

This comment captures a widespread sentiment that is frequently articulated in India: the concept of a unifying and unified Indian English *Dachsprache* is rejected in favour of preponderant "sub-national sentiment and sub-national competition" (Austin 1966: 306). More than thirty years after Austin made his comment, Krishnaswamy & Burde go even further and allege that the concept 'Indian English' defies all scholarly endeavour (1998: 4):

> In spite of some research, writing and thinking in the area of 'Indian English', there is no comprehensive perspective or clarity on the status and meaning of English in the wonder that is India. As in many other areas, in the land of contradictions there are only questions awaiting answers.

Other scholars do not share their epistemological pessimism and continue to assert that there is indeed an entity "Indian English" Sedlatschek (2009) and

1. Kachru's importance for the evolution of the study of New Englishes will be the subject of Chapter 2.1.2.

Mukherjee (2007, 2010) are recent cases in point. However, even though the bulk of Mukherjee's earlier article is devoted to the treatment of Indian English as a stable autonomous system, he still sounds a cautious note: "I firmly believe that English will always remain secondary in processes of Indian identity construction" (2007: 174).

This position supports Schneider's characterization of Indian English as hovering uneasily between stage 3 (nativization) and stage 4 (endonormative stabilization) of his Dynamic Model depicting the evolution of Postcolonial Englishes:[2]

> Given the current situation, Indian English is likely to stay and to defend the compartmentalized domains which it controls, and probably to keep growing further in terms of numbers and competence levels. At present it seems unlikely, however, that the language is going to cross the line and acquire new, emotionally more laden functions in Indian society. (Schneider 2007b: 172f.)

To take stock of what has been said so far: scholars writing both from inside and outside India unanimously agree that English in India is not a national language in the sense that it "has not adopted the function of an identity-carrier", neither does it "signal a pan-Indian identity" (Schneider 2007: 167). In Schneider's terminology, the final stage 5 of his model, namely dialect differentiation, is not feasible yet for Indian English because the socio-political parameters for the achievement of this final stage have not been set. Vaish's reference above to "at least 18 different varieties that are the regional standards of India" should then not be taken as an indicator that dialect differentiation has already set in and that stage 5 has been reached, after all. On the contrary, her comment shows that when it comes to Indian English, there is a curious reluctance to try and see the wood for the trees.

The perspective from language pedagogy is not identical with an explicitly political, anti- or post-colonial view, but frequently coextensive with it. Teachers of English in India, especially those outside the metropolitan areas, stress the unequal access to English within the Indian education system (cf. Sheorey & Bhaskaran Nayar 2002, Verma 2002). This observation translates into the claim that English in India is an elitist language, a resource that is employed by a powerful minority to assert and maintain their dominance in society (e. g. Annamalai 2004). The journalist S. Anand (1999) is even more outspoken in this respect: he avoids the euphemism *rural vs. metropolitan*, which is frequently used to gloss over the principally caste-based inequalities in the Indian educational system and in society at large. According to him, the debate about Indian English is simply irrelevant for the vast majority of Indians, for whom English is plainly a foreign language without any further qualifications. The question whether Indian English is a native or

2. Schneider's Dynamic Model will be given more space in Section 2.1.4.

a non-native variety remains purely academic except for a tiny Brahmin elite, who squibble with the white man to beat him at his own game (Anand 1999).[3]

Taking account of the frame of reference for English in India as sketched above, we might twist Benedict Anderson's famous phrase around and characterize Indian English as an imagination in search of a (speech) community.[4] Even though a wide range of studies on individual aspects of Indian English phonology, lexicon and syntax are available by now, this work has so far not culminated in a comprehensive grammar of Indian English. Moreover, the mismatch between the actual size of the Indian English speech community and the scholarly activity directed at the study of IndE is striking: Schneider (2007a: 352) reports that only nine out of 64 papers published by *English World-Wide* over the last five years deal with English in Asia (as e.g. compared to 14 contributions on English in Africa). A closer look at these nine papers then reveals that exactly two of these were on Indian English, just as many as on Singaporean English.

Indian English remains quite literally conspicuous by its absence: the most accomplished achievement in the field to date, the massive *Handbook of Varieties of English* (Kortmann et al. 2004), contains a mere sketch of some IndE syntactic features that does not even follow the general format for the syntactic descriptions of varieties which otherwise appear in the *Handbook*.[5] What is worse, IndE and IndE features are not included in the *Handbook's* "Global Synopsis: morphological and syntactic variation in English" (Kortmann & Szmrecsanyi 2004). Moreover, even though the Indian English component of the *International Corpus of English* (ICE-India) has been around since 2002, this rich database features mainly in cross-sectional studies of e.g. article use in New Englishes (Sand 2004) or particle verbs in New Englishes (Schneider 2004). So far, no study has tried to take stock of the syntax of spoken *standard* Indian English as exemplified by the data collected in ICE-India.[6] Such a corpus-based account of IndE usage that moves away from impressionistic notes on IndE "peculiarities" would be a prerequisite for any further considerations of the question whether we are dealing with "English in India" or "Indian English", now or in the near future.

3. The native-non-native distinction and its relevance for the study of New Englishes will be investigated more fully in Section 2.2.

4. Cf. Anderson (2006).

5. IndE has fortunately been given a more thorough treatment in the new *Electronic World Atlas of Vareties of English* (www.ewave-atlas.org).

6. Sedlatschek (2009) is based on a purpose-built corpus.

1.2 Aims and scope of this study

My study is an attempt to partly fill this gap by investigating some aspects of the syntax of spoken IndE on the basis of the conversation files of ICE-India. The speakers providing the data have been chosen to represent the standard usage of their respective variety in accordance with the general ICE-corpus design (cf. Greenbaum 1996). Thus, the picture that will emerge from the study can be taken to be representative of educated spoken Indian English. My justification for focussing exclusively on spoken language and further on spontaneous dialogue is twofold: first of all, it is a sociolinguistic truism that naturally occurring conversation is as close as we can possibly get to the vernacular, the variety unfettered from the prescriptive norms and standards of writing. We might assume that for IndE as well as for New Englishes generally, the written norm is still more or less exonormative, modelled on British English or International English usage. The point of departure for nativization and endonormative stabilization, to use Schneider's terms again, is thus necessarily speakers' actual language use in interaction with other members of their speech community. Further support for concentrating on dialogue data comes from Matras & Sakel (2007: 847–848):

> any type of language change will begin at the level of the individual utterance in discourse. [...] In order to understand the triggers behind various mechanisms of change in situations of contact, we must therefore explore multilingual speakers' motivations to adopt ad hoc solutions and strategies in response to the communicative challenges that face them when structuring individual utterances in discourse.[7]

This study, then, will attempt to assess some structural properties of the IndE vernacular, with the aim of contributing to a descriptive account of the range of variation pertaining to the norms of spoken Indian English. However, the apparent contradiction between using a corpus of *standard* IndE to get at the *vernacular* is one which is bound to arise in multilingual societies where English is an official language, but not necessarily speakers' first language. As will be discussed in more detail below (cf. Chapter 2), the study of New Englishes in general and IndE in particular has been bedevilled by many ideologically charged debates, prime among them the "native speaker-debate". A corollary of the native-non-native distinction is the classification of innovations in the New Englishes as mistakes, indicative of fossilization in the process of second language acquisition: speakers of New Englishes have just missed the target (language).[8] Singh's (2007) dismissal

7. The theoretical ramifications of Matras & Sakel's hypotheses will be considered in more detail in Section 2.3.

8. For an elaboration and criticism of this view see Bhatt (2002: 84–86).

of this position captures the tension between *standard* and *vernacular* with respect to Indian English:

> speakers (=/ learners) of Indian English are native speakers of Indian English just as speakers of Midwestern American English are native speakers of Midwestern American English [...]. Indian English needs to be studied with all the opposites of standard British and American English, Englishes that are treated as marginal by speakers and sociolinguists of standard British and American English. [...] The only thing to remember is that we are talking about speakers and NOT learners. (Singh 2007: 43)

While it is clearly undeniable that English is a foreign language for the vast majority of Indians, there is a significant community of "native speakers/users" of Indian English (cf. Singh 2007: 38), and it is precisely this speech community which is represented in ICE-India. To put it differently: the speakers who provided the ICE-India conversation files are proficient speakers of IndE in the sense of Mufwene (1998: 117):

> the umbrella term for the arbiter of well-formedness and appropriateness in any community is the proficient speaker, one who is fully competent in a language variety according to the established norm of the community using it. Whether or not such a speaker is typically native will vary from one community to another.

The ICE-India data, then, give us access to IndE usage according to the "established norm of the community". They further allow us a glimpse of the vibrant multilingual contact scenario in which speakers of IndE are immersed, a point that is vital for the present study.

One of the aims of this study is to describe such "established norm[s]" for spoken IndE within a specific domain, namely that of "word order and discourse organization", a category applied in Kortmann et al. (2004). I will investigate a cluster of linguistic features that speakers have at their disposal to express an array of discourse-pragmatic meanings, for example structuring their propositions, foregrounding or backgrounding salient information, expressing their attitude toward the contributions of others in the ongoing discourse. From this perspective, the domain of discourse organization and discourse management comprises two subdomains: the expression of information structure by non-canonical word order and the domain of utterance modifiers, for example discourse markers.[9] The linguistic features belonging to this domain rank high on impressionistic lists designed to capture features that are supposed to be "typical" of Indian English (e.g. McArthur 2003: 323–324, Kachru 1986: 40, Sailaja 2009: 53–59), but have so far not been treated in greater detail, much less from a unified perspective.

9. The term "utterance modifiers" is derived from Matras (1998), cf. Chapters 2.3.2 and 4.6 for a fuller treatment.

More specifically, I will investigate the following features. First of all, I will focus on the non-initial existential *there*-constructions (e.g. *Food is there*, cf. Nihalani et. al. 1979: 177) which are not attested in any other variety of English. Next, both topicalization (e.g. *Preparation of the marriage I'm talking about* (ICE-IND: S1A-095#165)) and left dislocation involving resumptive pronouns (e.g. *Rajasthan people **they** eat lot of pappad*[10] (ICE-IND: S1A-008#94)) have been observed to occur at a higher frequency in the New Englishes. I further treat cleft constructions in passing to complete the overview of syntactic features in the domain of discourse organization.

Within the domain of utterance modifiers, I focus on two expressions that are almost stereotypically associated with IndE. I will have a closer look at the distribution and function of the invariant tags *isn't it* and *na/no* (cf. Nihalani et al. 1979: 104), and further at the focus markers *only* and *itself* (e.g. *He came here today only* (1979: 132); *Can I come and see you today itself?* (1979: 105)).

Since ICE-India provides social and linguistic information about each speaker, it is possible to correlate linguistic variation with independent variables such as age, sex, education, mother tongue. Such an analysis will then allow tentative conclusions about the future of each structural feature under consideration: if, say, an expression is already firmly entrenched in the speech community, it is more likely to eventually become part of an Indian English endonormative standard than an expression with a less general distribution, e.g. an expression which is restricted to speakers of a specific age and/or region. This is in line with Sridhar's characterization of a future South Asian English

> standard or acrolect that (1) is not too strongly marked by varietal features of one particular region, (2) is free from stigmatized features such as gross agreement violations, and (3) enjoys pan-South Asian distribution, intelligibility, and positive evaluation. (1996: 67)

The domain of discourse organization comprises structural features that are widely attested for New Englishes generally (e. g. topicalization, invariant tags) as well as constructions that are unique to IndE (or Indian Englishes, if we subsume South African Indian English under that label), such as the non-initial existential *there*-construction and presentational focus markers *only/itself*. Four main explanatory parameters are generally invoked in the relevant literature when it comes to innovations in the New Englishes: (a) substrate influence, (b) retention of earlier, non-standard features from the superstrate, (c) universals of second language acquisition, (d) mechanisms akin to pidginization and creolization.[11]

10. *Pappad*: thin, crispy, spicy bread.

11. Cf. e.g. Mesthrie (1992: 152–221) for a thorough discussion of (c) and (d).

I would argue that referring to "substrate influence" merely provides a label, but is nowhere near an explanation. First of all, language contact between Indo-Aryan and Dravidian languages in South Asia dates back thousands of years, with considerable convergence as a result.[12] Many syntactic features (e.g. enclitic focus markers) are widespread in both language families, so that it becomes almost impossible to identify exactly the language which served as the actual source language. Moreover, language contact is not only an historical fact, but an ongoing process in South Asia where multilingualism is the norm rather than the exception. Hindi and English are the national link languages or "pragmatically dominant languages" (cf. Matras 1998, 2009) in India, so to speak, with potential for influence or interference with all other Indian languages.

Another problem with the label "substrate influence" is related to what purports to be its outcome: the label is indiscriminately applied to borrowings (from lexical items and morphemes to syntactic categories and structures), calques, and the many other ways in which speakers currently conversing in a target language can make use of a more or less abstract concept derived from a source language. If we subsume all these phenomena under one convenient label, we reduce the dynamics of continuous language contact to something remote and static in the past. We also deprive ourselves of a better understanding of what actually happens in contact-induced change in context: if we indeed take for granted that every language change starts off as a "process of innovation by one or more individuals" (Thomason 2008: 51), then we should also shift our perception of speakers in multilingual contexts who

> are not only receivers and imperfect learners but also creators who use what they find in one language and sociocultural environment to shape another language in novel ways – they do not simply imitate grammatical categories, or produce imperfect copies of such categories; rather, they are likely to develop new use patterns and new categories on the model of other languages. (Heine & Kuteva 2005: 37)

Developing new use patterns on the model of other languages may take a variety of forms, all of them realizations of communicative strategies available to the multilingual speaker. He or she is neither a "slave" of his/her substrate language nor a passive recipient of some version of the target language; rather, the multilingual speaker has an array of stylistic choices at his/her disposal which reflect the interaction between the languages available to the speaker in different ways and to different degrees.

Explanatory parameter (b), retention of earlier non-standard features or superstrate influence, has been highlighted by Mesthrie (2006b: 277) as a possibility that is frequently overlooked in studies of New Englishes:

---

12. A brief overview of *sprachbund* phenomena in South Asia will be given in Section 3.2.2.

> the "shape" of the superstrate should not be taken for granted. [...] the notion of a target language (TL) is an idealization: more often, and certainly outside the classroom, the TL was a varied and "moving" target.

Mesthrie's argument "for a less substratophile interpretation of New Englishes variability" (2006b: 279) may take two forms: as he himself stressed in a series of related articles (Mesthrie 2003, 2006a, 2006b), the first speakers and/or teachers of English coming into contact with a local population were frequently non-native speakers themselves or speakers of a non-standard variety of English.[13] Mesthrie points out that linguists would be well advised to first identify the variety of English that served as the target language in any given contact situation rather than to turn straight away to substratist explanations.

Another way of accounting for variability in New Englishes by exploring the input from the model language(s) was pioneered by Sharma (2001) in her study on the pluperfect in Indian English. By drawing on data from parallel corpora of written British English, American English, and Indian English, Sharma was able to show that the categories of tense and aspect display considerable variability in British and American English. She stresses "the importance of taking into account variability and ambiguity in the lexifier or native variety itself when considering possible sources of new form–function relationships in a new variety" (Sharma 2001: 366). In the context of this study, Mesthrie's diachronic superstratist strategy is unlikely to play a role; however, Sharma's synchronic approach to variable superstrate input and her pronouncement that "ambiguity in native English usage represents an important trigger for variability and change" (2001: 370) should be kept in mind.

Explanatory parameter (c), namely processes universal to second language acquisition, is of course a prime suspect for many innovative features that are found in more than one of the New Englishes:

> the large number of similarities across L2 Englishes [...] needs to be explained more carefully than in the past, where the default assumption has often been interference from the substrates. Since there are over a thousand of these substrate languages in Africa-Asia, the explanation of interference has to be considerably fine-tuned. It is *prima facie* implausible, areal linguistics notwithstanding, that over a thousand languages should induce the very same (or very similar) influences. This would be tantamount to claiming that all the languages of Africa-Asia are the same in structure, united in their difference from English. (Mesthrie 2008: 634)

13. The notions "standard/non-standard" should be taken with a grain of salt here, as the global expansion of English began before the standardization and codification of the language was anywhere near its end. The terms "uneducated" or "non-literary" English are perhaps more appropriate in this context.

It is then "the psycholinguistics of second-language processing of a cognitive system like English" (*ibid*.) which should be studied to account for *angloversals*, a label given to innovative features which are found across a wide range of New Englishes (cf. Mair 2003).[14] Again, the realm of discourse organization as outlined above, with some features unique to IndE and some that are common in many other New Englishes, will prove promising for asking a principled question about contact-induced variability and change: why should some areas of grammar be more (or less) susceptible to contact-induced language change than others? Why should, say, different word order patterns be amenable to different explanatory parameters? To take one example: existential sentences with clause-final *there* are unique to IndE, whereas topicalization is very common in the New Englishes generally. Why should speakers resort to an expressive strategy which is highly variety-specific in one case, but not in the other? To put it differently, why do the resources and strategies that are available to speakers in multilingual contexts converge in some cases and produce similar results, but diverge in others? In order to get closer to an answer to these questions, I will analyze the form, frequency, distribution and discourse function of each individual feature in detail, paying close attention to the discourse strategies they may serve within multilingual communicative contexts.

1.3 Structure

In order to address the questions outlined above, this study will proceed as follows. Chapter 2 is devoted to the conceptual background informing all subsequent chapters and will consider some theoretical concepts relevant for this study as they have evolved together with research on New Englishes. The very term "New Englishes" has come under increasing scrutiny in recent years, as Section 2.1 will show. Section 2.2 focuses on the theoretical and ideological issues related to the fundamental concept of "native speaker", while Section 2.3 reviews current research on contact-induced language change. Section 2.4 disentangles the related, but by no means identical notions of "norms" and "standard", notions which repeatedly came to the fore in the two preceding sections. Taken together, the four sections will trace the changing concerns and research strategies of the field up to the present.

Chapter 3 will delineate the communicative space in India and the changing position of English within that space. I prefer the notion "communicative space"

14. This rather catchy notion has so far not received much empirical backing, cf. e.g. Hundt (2007).

(which I take from Oesterreicher 2001) to the widely used term "linguistic ecology" (cf. Mufwene 2001). Mufwene's concept is closely linked to his "Founder Principle", i.e. the original contact scenario as the main determinant of the resulting varieties' status and structure. While this is certainly highly relevant for assessing broad macro-linguistic lines of development for a particular variety, the Founder Principle is too blunt an instrument for a more detailed analysis of actual linguistic structures. "Communicative space", on the other hand, has the advantage of encompassing not only the multilingual scenario, but further takes the parameters orality-literacy into account.

Chapter 3.2 therefore presents a brief typological survey of the Indian languages with particular emphasis on *sprachbund* phenomena apparent in the Indo-Aryan and Dravidian languages. These typological considerations will be helpful in contextualising possible substratist explanations in later chapters. The Chapters 3.2.3 and 3.2.4 will extend the notion of *sprachbund* to include convergence phenomena that are pragmatic rather than phonological or syntactic, the traditional preoccupations of areal typology. Chapter 3.3 will touch upon some aspects of the dynamics of the Indian communicative space and the position of English within it.

Chapter 4, then, is devoted to the corpus-based analysis of the domain of discourse organisation in spoken Indian English. Sections 4.1.2 to 4.1.4 give an assessment of the data used for this study, namely the "private dialogue" files of ICE-India (S1A 1–100). The general format of the ICE-project is too well known to be repeated here; but since conditions for collecting and processing data for the individual ICE-corpora necessarily vary from country to country, it makes sense to consider the ICE-India data in more detail. Obviously, the input to any corpus-based project crucially determines and also constrains the generalizations to be made that are based on the analysis of the data.

Each of the following subchapters devoted to discourse-pragmatic features of IndE will include an analysis in terms of quantificational sociolinguistics: are there discernible patterns of variation along the external variables sex, age, educational level, mother tongue? Do such patterns hold for more than one of the features? But even though this part of the investigation will be highly informative when it comes to the diffusion of features throughout the IndE speech community, it is clearly not sufficient, for both theoretical and conceptual reasons. Quantifying features undoubtedly has its merits, but it also neglects the interactive multilingual contexts in which these features are actually made meaningful. As said above, the multilingual speaker is not a "carrier" of features whose occurrence is largely predetermined by age, sex, etc.; rather, he or she makes choices within the immediate communicative situation which, in turn, is placed within a specific communicative space. Such choices may include borrowing, calquing, or using a specific construction in

contexts which go beyond the original syntactic or discourse-pragmatic constraints that were originally associated with it. The pertinent question here is whether we can reconstruct speakers' cognitive processes in choosing the forms they did, and if these cognitive processes in turn tell us something about borrowability in language contact situations.

Chapter 5 is an attempt to summarize and generalise the findings from the preceding chapter, and to project them into the future. My aim here is twofold: first of all, I turn to the theoretical implications of my findings, trying to asses which explanatory parameters outlined above account best for speakers' choices in multilingual settings. Speakers' individual solutions may give rise to new usage patterns which in turn, under specific conditions, may spread throughout the speech community: in a final step, I will attempt to evaluate whether the features identified for spoken IndE will eventually become part of a future standard Indian English.

CHAPTER 2

Conceptual background

The aim of this chapter is to spell out the theoretical assumptions and methodological choices which inform the present study. In doing so, it touches upon some of the most pertinent theoretical concepts that have developed along with the study of New Englishes. This chapter is therefore partly a history of ideas in the field, ideas which were paramount in advancing research on the topic but which have, perhaps inevitably, come under increasing pressure from alternative approaches.

As stated in the preceding chapter, this study gives pride of place to the multilingual Indian speaker of English, examining the communicative strategies which arise naturally in multilingual verbal interaction. It thus draws on current concepts in the field of contact linguistics, where a host of new perspectives, theories and methodologies for studying the outcome(s) of language contact situations have been advanced in recent years. It was possibly Thomason and Kaufman's *Language Contact, Creolization, and Genetic Linguistics*, published in 1988, which sparked the renewed interest in motivations for and particularly constraints on contact-induced language change – "renewed" because Weinreich's 1953 monograph *Languages in Contact* has to be given due credit, even though it stood more or less alone in its time. To the contemporary student of New Englishes, it may appear somewhat odd that the tools of contact linguistics are only now being applied to their study – after all, New Englishes are contact languages *par excellence*. Section 2.1, a short historical tour through the field of New Englishes, and concomitantly Section 2.2 on the concept of "native speaker", provide a partial solution to this puzzle: it simply took some decades before New Englishes gained recognition as varieties in their own right, rather than as fossilized stages of imperfect language learning. Section 2.3 briefly introduces some of the notions of contemporary contact linguistics that are relevant for the study of New Englishes and that feature prominently in the present study. Finally, Section 2.4 focuses on models of standardization and their application to the New Englishes.

2.1 New Englishes and outer circles

If it is a sign of maturation for a (sub-)discipline that its original assumptions, the key distinctions it fought to establish, are slowly being discarded in favour of

a more integrated approach, then the study of New Englishes has certainly reached its prime. Platt et al. (1984) were the first to use the term "New Englishes" and to provide the now well-known definition as well as a broad descriptive overview of typical linguistic features. Braj B. Kachru, on the other hand, made the case for the marginalized Englishes by drawing attention to the fact that some Englishes are more equal than others depending on their degree of nativeness, i.e. the Inner Circle varieties. Now, in the first decade of the 21st century, both the theoretical landscape and the sociolinguistic profiles of countries where New Englishes are spoken have changed considerably, and it is the aim of this chapter to trace the development of those ideas and concepts that originally demarcated the field.

2.1.1 Ideology and terminology: From *English* to *Englishes*

The early eighties of the last century saw a considerable rise of interest in varieties of English, or *Englishes*, around the world. The first journal devoted to the new field (appearing in West Germany), however, was titled *English World-Wide* rather than *English**es** World-Wide*: its founder Manfred Görlach "considered choosing *Englishes* as its title but refrained from doing so because the plural form was still felt to be unacceptable" (Schneider 2007b: 3). With their monograph *The New Englishes* (1984), Platt, Weber and Ho provided the budding field with a new label as well as a working definition:

> We shall consider that a New English is one which fulfills the following criteria:
>
> 1. It has developed through the education system. This means that it has been taught as subject and, in many cases, also used as a medium of instruction in regions where languages other than English were the main language. [...]
> 2. It has developed in an area where a native variety of English was not the language spoken by most of the population. For various reasons, which we will discuss later, pidgin and creole languages are not considered to be native varieties of English.
> 3. It is used for a range of functions among those who speak or write it in the region where it is used. This means that the new variety is used for at least some purposes such as: in letter writing, in the writing of literature, in parliament, in communication between the government and the people, in the media and sometimes for spoken communication between friends and the family. It may be used as a lingua franca, a general language of communication, among those speaking different native languages or, in some cases, even among those who speak the same native language but use English because it is felt to be more appropriate for certain purposes.

> 4. It has become 'localized' or 'nativized' by adopting some language features of its own, such as sounds, intonation patterns, sentence structures, words, expressions. Usually it has also developed some different rules for using language in communication. (Platt et al. 1984: 2f.)

Other scholars objected to the term: "the label *new* reflected primarily a shift of attention in western, Anglocentric scholarship" (Schneider 2007b: 3). *New Englishes* were, after all, far from 'new' for their speakers, and some extraterritorial varieties of English that are subsumed under the label (such as Indian English) are actually much older than, for example, Australian English or New Zealand English.

One solution to remove the hierarchy as well as the bias that was felt to be inherent in the term was to coin another, more inclusive one: *World Englishes* became the programmatic title of another important journal serving the field, giving all varieties of English equal status and refusing to privilege the Western point of view.

A further point of criticism directed at the new label concerned the delineation of the category *New Englishes*. Platt et al.'s definition of "New English" as quoted above explicitly excludes Pidgins and Creoles from the category "native variety of English". For scholars such as Mufwene (e.g. Mufwene 2001), assigning Pidgins and Creoles to a class of their own constitutes another case of "the fallacy of creole exceptionalism" (DeGraff 2005), that is the special position given to Pidgins and Creoles in linguistic theory. Pidgins and Creoles were treated separately from other New Englishes because creolization was taken to be a process quite unlike any other "natural" process of language evolution. According to Mufwene, the reason for assigning Pidgins and Creoles to a class of their own lies, however, elsewhere:

> The main implicit criterion, which is embarrassing for linguistics but has not been discussed, is the ethnicity of their speakers. Most hypotheses proposed in creolistics to account for the development of creoles would have been better thought out, had it not been partly for this factor, as strong as my accusation may sound. (2001: xiii)

Schneider's term *Postcolonial English* tries to steer clear of such ideological debates; the term is further intended to emphasize that the (post-)colonial contact situation is, after all, crucial for his theoretical framework:

> the varieties under discussion are products of a specific evolutionary process tied directly to their colonial and postcolonial history. I am concerned with developmental phenomena characteristic of colonial and the early phases of postcolonial histories until the maturation and separation of these dialects as newly recognized and self-contained varieties; hence, the term is taken to encompass all forms of English resulting and emerging from such backgrounds. (2007b: 3)

Schneider's "Dynamic Model of the evolution of Postcolonial Englishes" has largely superseded Kachru's descriptive and explanatory parameters; but before I turn to a brief consideration of the Dynamic Model, I will first remain true to the chronological development of the field and discuss the impact of Braj Kachru's work.

2.1.2 The legacy of Braj Kachru

Kachru's work on Indian English dates back to the sixties; his project of gaining recognition for the New Englishes became more vociferous in the eighties and early nineties, when he locked horns with Randolph Quirk several times over the issue of standards for English as a world language (cf. Quirk 1985, 1990; Kachru et al. 1985, 1991).[1] Kachru visualized the global spread of English with his model of the three concentric circles as in Figure 2.1.

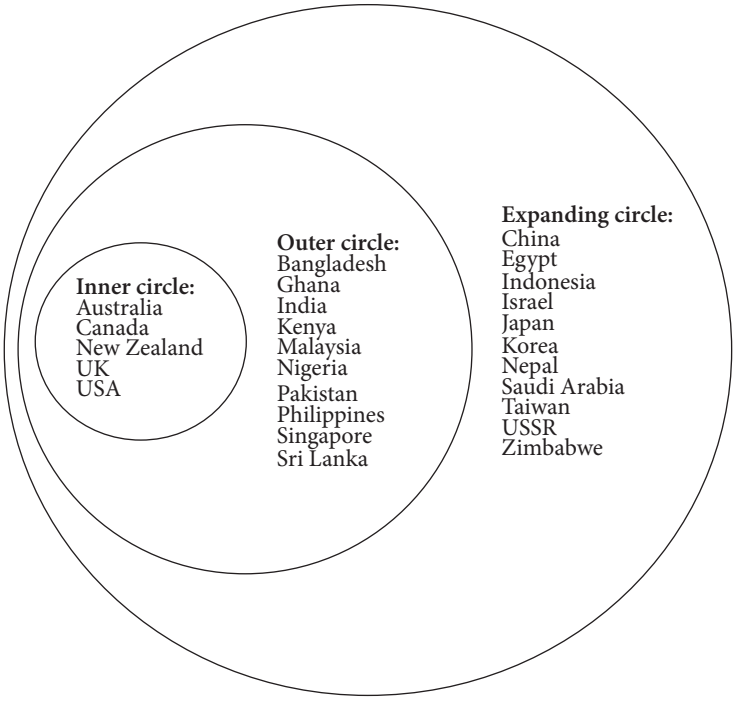

Figure 2.1. Braj Kachru's model of Inner Circle and Outer Circle varieties of English (in Bauer 2002: 23)

1. Kachru 1983, 1986 are collections of earlier papers, which frequently make similar points.

Inner Circle varieties are the L1 varieties in Western, (mainly) monolingual countries such as the UK, the US, Australia and New Zealand. The Outer Circle comprises many Asian and African countries, where English used to be the colonial language and where it has kept its place within a multilingual scenario for a range of functions. The Expanding Circle, finally, comprises countries where English is taught as a foreign language. Kachru was relentless in pointing out that the former unequal relationship between colonizers and colonized was still apparent in the attitudes of speakers, teachers and linguists from the Inner Circle towards English(es) in the Outer Circle. The very fact of the globalization of English entailed not only diversity, but also pluricentricity: the "Eurocentric, Judeo-Christian ethos of the language" (Kachru 1988: 3) was challenged by localized varieties "reflecting the non-Western identities" (*ibid*.) of their speakers. This challenge was perceived as a challenge to standard English by scholars like Quirk (1990), who maintained a categorial distinction between native and non-native varieties of English. He saw Outer Circle Englishes generally on a par with English in the Expanding Circle, i.e. English as a learner variety:

> The problem with varieties in this branch is that they are inherently unstable, ranged along a qualitative cline, with each speaker seeking to move to a point where the varietal characteristics reach vanishing point, and where thus, ironically, each variety is best manifest in those who by commonsense measures speak it worst. (Quirk 1990: 5f.)

Kachru reacted to Quirk's accusation of an "anything goes" attitude in linguistics and language teaching and in turn blamed Quirk for espousing "deficit linguistics" (Kachru 1991: 4). Outer Circle varieties typically develop in multilingual contexts and are thus subject to what Platt et al. termed "nativisation" and what Kachru refers to as the "multilinguals' creativity" which "has introduced a wide range of innovations, linguistic, literary and sociocultural, which are not within the canon of the Inner Circle" (Kachru 1988: 6). Far from representing "locally acquired deviation from the standard language" (Quirk 1990: 8), innovative IndE usages are a natural development in an Outer Circle speech community that is characterized by functional, if not genetic nativeness (cf. Kachru 1998).[2]

Much of Kachru's work on IndE has been devoted to characterizing the "Indianness" of IndE along these lines, and it is this aspect of his work that has subsequently attracted criticism. However, contemporary critics may do well to keep in mind the theoretical and historical context in which his ideas developed:

2. Cf. further Section 2.2.

> [W]hat a long way Kachru had to travel to even begin to develop an account of Indian English free from the overt acceptance of colonial control! His achievements in the implementation of Dustoor's post-colonial programme, of getting the sahibs to agree that the varieties of English used in the ex-colonies should count as varieties of English (though admittedly marginal, 'non-native' varieties) subject to the normal laws of naturalist linguistic description, have to be seen to be believed. (Dasgupta 1993: 123)[3]

Despite this initial praise, Dasgupta himself is one of the fiercest critics of what has been termed "the Kachruvian enterprise", as the following section will show.

2.1.3 "The Kachru catch"[4]

Over the years, criticism of Kachru's work has accumulated. Kachru's critics, who by now seem to be in the majority, agree that what we know from Kachru's work about IndE is largely unreliable, for empirical and conceptual reasons. Kachru's critics have noted early on that he draws most of his examples of "Indianisms" from Indian literature written in English:

> Since most of Kachru's examples of Indianisms are from creative writing in English done by exceptionally competent writers who created those expressions for special literary effects, it can be argued that they are not features of IE as it is used by educated bilinguals in social situations. (Parasher 1991: 62)

Parasher's comment notes that Kachru's description of IndE is necessarily limited in scope due to its limited database: only a highly specific written register is taken into account. More damaging is the criticism made by Dasgupta who questions Kachru's methodology on principle:

> Kachru's contextualizations do not, because they cannot, instantiate the workings of a serious socio-linguistic analysis whose details would address a specialized and rigorous audience. We are not looking here at mere failures of implementation irrelevant to the intrinsically acceptable programme Kachru had set out to implement. Rather, we are looking at omissions and inaccuracies which reflect the fact that precision and rigour were unavailable to him on principle. His very programme handicapped him. The item-to-item mini-stories he set out to tell could not have been narrated more satisfactorily on any systematic basis, for there was no thread that could have linked them to any system and given his overall description more firmness. (Dasgupta 1993: 128)

---

3. P. E. Dustoor was one of Kachru's teachers and one of the first linguists to promote the notion of "Indian English".

4. "The Kachru catch" is the title of the section in Krishnaswamy & Burde (1998: 30–36) devoted to dismantling Kachru's terminology and methodology.

In short, Kachru's descriptive account of IndE, exemplified by, for example, Kachru (1994), falls short of being explanatorily adequate because it merely represents an unstructured and decontextualized, even pre-theoretical list of features. As such, his approach is effectively detrimental to any further in-depth exploration, as Sedlatschek has observed:[5]

> Generally speaking, in order to describe IndE as a variety in its own right, it is necessary to go beyond the simple observation of a phenomenon by illustrating the inroads of nativization with the help of a clearly defined set of criteria. Descriptions like Kachru's (1994) that are based predominantly on enumerations of linguistic features do not spell out those criteria and therefore run danger of creating a stereotypical picture of IndE – a picture that (unintentionally) turns out to be detrimental to an understanding of the variability and dynamics of IndE which make IndE 'uniquely Indian'. (2007: 28f.)

Kachru's original concept of the three circles of English has also been criticized: Bruthiaux (2003: 161) argues that "the Three Circles model is a 20th century construct that has outlived its usefulness", one of the reasons being that

> the Three Circles concept is a nation-based model that draws on historical events which only partly correlate with current sociolinguistic data, an inherent lack of theoretical consistency that goes back to Kachru's early articulation of the model (1984) through later refinements (1985, 1989). The result of this ambivalence regarding the nature and scope of the model – varieties, countries, speakers, or all three? – is that any explanatory power is lost as the analysis shifts from circle to circle. (2003: 172)

Bruthiaux concludes his critical evaluation of the Three Circles model by bidding the model farewell, but not without acknowledging its value in its time: "As incomplete, unfair, and outdated characterizations of varieties of English fade away, thanks in part to the Three Circles model, so ultimately must the model itself" (2003: 175).

What remains, then, as Kachru's outstanding achievement? Even his fiercest critics will concede that his work, more than that of any other scholar, was instrumental in putting the New Englishes firmly on the linguistic map. His contributions have triggered countless discussions about the status of New Englishes, both political and linguistic. IndE as well as other New Englishes are now considered varieties in their own right, a point of view which seems to be eminently commonsense now, but which was far from being a majority option (or even an option) when Kachru started writing. He further introduced the notion 'cline of bilingualism' to capture degrees of variation according to speakers' proficiency, and he

5. This passage is toned down considerably in Sedlatschek (2009: 28–30).

established the concept of functional as against genetic nativeness to better account for multilingual settings. These last points will be taken up in the next section, in the discussion of 'the native speaker'.

2.1.4 The Dynamic Model

Schneider's Dynamic Model of the evolution of Postcolonial Englishes (PCEs) has already been mentioned as the most attractive and authoritative successor to date to Kachru's model of the three Circles. The Dynamic Model proceeds from the following basic assumption:

> Fundamentally, the evolution of PCEs is understood as a sequence of characteristic stages of identity rewritings and associated linguistic changes affecting the parties involved in a colonial-contact setting. (Schneider 2007b: 29)

Crucially, such a perspective abandons the differentiation made by Platt et al. (1984) between New Englishes and Pidgins and Creoles, and thus answers to the criticism levelled at the linguistic community for according Pidgins and Creoles an unwarranted special status. The model is designed to accommodate the development of all varieties of English that can be traced back to a colonial contact scenario, be it Irish English or Jamaican Creole. After all, "the parties involved in a colonial-contact setting" mentioned in the quote above are basically the same: the colonizers and the colonized, or the English-speaking settlers and the non-English-speaking indigenous population (the "STL strand" and the "IDG strand" in Schneider's terminology), whose "identity rewritings" drive the micro- and macrolinguistic development of the newly evolving variety of English. In this perspective, the distinction 'native/non-native' roughly corresponds to the original distinction between settlers and indigenous population; however, the distinction is neither bound to a nation state nor as static as it appears in Kachru's model:

> I wish to apply my notions of the STL and IDG strand to speech communities, frequently defined along ethnic lines, as agents in an ongoing process. [...] both groups who share a piece of land increasingly share a common language experience and communication ethnography. (Schneider 2007b: 32)

To conceive of the evolution of PCEs as an "identity-driven process of language convergence" (2007b: 30), then, shifts the focus towards the agentivity and interaction of both speech communities over time, two dimensions that were more or less absent from Kachru's model.

Schneider stipulates five stages for the development of PCEs (Schneider 2007b: 29–55). The first stage, "Foundation", unsurprisingly involves the arrival of English-speaking settlers in a new territory. For India, the initial foundation phase

is frequently associated with a precise date, namely the last day of December 1600, when Queen Elizabeth I granted a royal charter to the East India Company to pursue trade in South and Southeast Asia (Schneider 2007b: 162, Kachru 1994: 502). However, the first Englishman to set foot on Indian soil in the name of the East India Company did so only in 1612, at Surat in what is now coastal Gujarat, and by the end of the 17th century the Company had only established a handful of trading posts in coastal areas. English was thus no more than another traders' language for an extended period of time, with little to no impact on the Indian communicative space.

Schneider points out that the onset of his second phase, "Exonormative stabilization", is difficult to pin down for India (2007b: 163); this phase typically involves the consolidation of colonial power in the new territory, with English serving as the language of administration, jurisdiction, and education (2007b: 36). Schneider dates phase two from 1757 to around 1905, a period in which the East India Company gained military and political control over the Indian subcontinent. In 1858, following the Indian Mutiny, British India came under direct Crown rule. Krishnaswamy & Burde take the date of the actual Mutiny (1857) as the cut-off point for their "dissemination phase" in the history of English in India, lasting up to 1904 (1998: 98).

Both Schneider and Krishnaswamy & Burde then agree that the next stage in their respective models (phase three, or "Nativization" in the Dynamic Model, "the institutionalization phase" for Krishnaswamy & Burde) begins well before India's independence in 1947, when English attains the status of a second language for a growing number of the IDG speech community. 'Nativization', a familiar term from Kachru's discussion of IndE properties, pertains to

> indigenization of language structure [...] the emerging new variety is gradually enriched with additional structural possibilities, and ultimately parts of its grammatical make-up (i. e. its lexicogrammatical constraints) are modified. Hence, grammatical nativization in PCEs typically start out with a specific set of patterns which appear to occur more frequently than others. (Schneider 2007b: 46)

Schneider stresses two general observations about the structural consequences of nativization: first of all, that differences between PCEs and "Old Englishes" tend to be of a quantitative rather than a qualitative nature, and secondly, that differences become manifest primarily at the level of lexicogrammar, e.g. in verb complementation patterns (cf. Olavarría de Ersson & Shaw 2003, Mukherjee & Hoffmann 2006, Schilk 2011), particle verbs (Schneider 2004, Sedlatschek 2009: 149–179), or collocations (Schilk 2006). I will return to the discussion about the projected outcomes of language contact situations in the next section.

Schneider's fourth stage, "Endonormative stabilization", requires not only political independence, but "cultural self-reliance" (2007b: 48) and typically gives rise to "a new, locally rooted linguistic self-confidence" (2007b: 49), that is, speakers of the former colonial language now claim the language as their own and consider themselves members of a speech community whose norms are negotiated by themselves. This self-confidence often becomes apparent in endeavours to codify the new variety without recourse to exonormative standards. It is at this stage that 'English in X' typically becomes 'X English' – but not necessarily so, as the Indian scenario demonstrates. After India's independence in 1947, English gained constitutional status as the Union's second official language and remained the language of administration, the legal system and of higher education, and the importance of English as the national link language has been growing steadily. However, it remains doubtful whether IndE has already reached stage four:

> As long as the *raison d'ĕtre* of Indian English is still essentially utilitarian and it is not a medium for community solidarity, the language is not likely to change its character and status in the near future. (Schneider 2007b: 173)

Mukherjee & Gries (2009) similarly describe IndE as moving towards stage four of the Dynamic Model; and Dasgupta, while fully endorsing the existence of an IndE speech community (cf. Dasgupta 1993: 118) characterizes the status of English in India thus:

> What does English in India mean and signify? This conjunctive question receives a conjunctive answer: English in India signifies technique and technology and technicality. It is not primarily a human language here. (1993: 214)

Dasgupta's skilful dialectics manages to assert the autonomy of the IndE speech community (vis-à-vis Quirk and others) and simultaneously to highlight the alienation that accompanies the use of English in India in his analysis. Naturally, Dasgupta has been challenged with regard to his dialectics, especially by d'Souza who explicitly rejects the notion that English in India is a metropolitan elitist language (cf. e.g. d'Souza 1997). However, for all claims to the contrary, Dasgupta's scenario seems to supply a clue for why "Indian English remains a complex, elusive, and problematic entity" (Schneider 2007b: 172): unless English in India becomes a "human language", that is, "a carrier of such a national identity and a language which is accessible to a major portion of society" (2007b: 171), it will not move further along the cycle.

Schneider's final and fifth stage, "Differentiation", involves the emergence of variation within the new variety:

> As a consequence of external stability, there is now room for internal differentiation: in the absence of an external challenge or need to demarcate a community against some outside entity, differences within a society and between individuals with respect to their economic status, social categories, and personal predilections come to light and can be given greater prominence. (2007b: 52–3)

In other words, once "English in X" has become "X English" by a general consensus of the speech community, the familiar processes of social stratification and social differentiation produce socially and culturally meaningful variation: the new variety is established as the reference variety or standard, and social or regional dialects acquire social significance with respect to the standard. Here lurks another paradox for New Englishes: endonormative standardization becomes an indispensable strategy of liberation from the "Old" Englishes, a linguistic declaration of independence:

> Without codification the varieties of English will continue to lack *locus standi*, not only in the eyes of the world, but also in the eyes of the speakers of the varieties. Without codification British English becomes the unnamed but ever present standard resulting in a learner mentality with New Variety English (NVE) speakers denied the freedom to use the language in new and innovative ways. If, when in doubt about a particular usage or point of grammar, an Indian speaker of English has to refer to exonormative materials, it inevitably undermines any kind of innovation that has gained ground within Indian English. (d'Souza 1999: 272)

Whereas standardization in the "Old" Englishes typically has the effect of blocking or at least slowing down linguistic innovation (cf. Milroy 2001), it may have the opposite effect for the New Englishes. I will consider the relation between norms and standards with specific reference to the New Englishes in Chapter 2.4. The intricate question of the shape of a future standard IndE will be the topic of the final chapter.

2.1.5 Outlook

Nobody who is interested in New Englishes can seriously deny the outstanding importance of Braj Kachru's work over the last decades. His model of the Three Circles of English provided the freshly hatched field with a welcome and much-needed framework for discussion. By necessity, these discussions went far beyond linguistic matters in a narrow sense, such as specific structural features of a particular New English; they always touched upon ideological biases that had gone unchallenged for too long – one particular ideological construct, namely that of the "native speaker", will be the topic of the next section. In such a growing and dynamic field, it is perhaps equally inevitable that new concepts supersede older models.

2.2 The *native speaker*: An elusive concept

"The native speaker" as a basic concept in linguistic theory has attracted considerable attention, specifically within the context of New Englishes, where the category has profound theoretical as well as political and economic implications and where these fields are difficult to keep separate. We have seen in the preceding section how Quirk – and, needless to say, many others – pitted the *native* speaker of English against learners of English, viz. the rest of the world. Quirk recognized two standards of native English for non-native speakers to aspire to, namely British and American English, and he explicitly evoked the danger of English becoming unintelligible if non-native varieties were encouraged to proliferate. As Wright unabashedly put it (2004: 11):

> We must unashamedly keep hold of a basic truth: native usage is by definition right. Other variants are not necessarily wrong, of course, but [...] the native is nevertheless still right.

Thanks to research on standardization processes and standard ideologies, the equation "native = standard" is at least no longer seen as a transitive relation: not every native speaker of English is also a speaker of standard English (even if he/she may be a writer of standard English, a distinction which tends to be overlooked). Kachru and others have further questioned whether a speaker of standard English is necessarily a native speaker of English, and it is the task of this section to explore this issue further.

2.2.1 The native speaker and the discourse of hegemony

In the vast literature on the native/non-native distinction, some major strands of argumentation can be distinguished. Many scholars since Kachru have been devoted to exploring the wider socio-political context of the distinction. One frequently held point of view stresses that privileging Inner Circle or native varieties of English over others ensures the continued global hegemony of the former imperialist powers:

> It is undoubtedly a truism that English has served, and still serves, as a prime cultural legitimation of the world division of labor into core and periphery. The dominant core ('center') has, over the last few decades now – in an effort to bolster and express its control over a greater linguistic diversity in the 'periphery' – attempted to "style" its English-use as definitive for the English-using community as a whole [...] This is precisely where the contemporary distinctions of native/non-native English appear – the core speakers are the standard bearers, but only those who have (had) the power to control the norm and successfully impose it on "others". [...] The standard serves only to reproduce socio-economic inequalities as it privileges only those who have access to its possession, leaving "others" disenfranchised. (Bhatt 2007: 55f.)

In that sense, continued adherence to the concept of the "native speaker" and to the norms that he/she provides is a serious obstacle to decolonization, both on the societal and the individual level. In maintaining cultural hegemony, there are also quite straightforward economic interests involved:

> There is, of course, an economic dimension which cannot be overlooked in any discussion of norms. Countries such as the US and Britain derive a significant income from exporting the English language, related products, and professional expertise. The economic power to be derived from exporting English provides the inner circle with incentives for maintaining control over both standards and attitudes. Quirk (1990: 7), for example, continued the argument for a global norm based on the inner circle, noting that it is "the leading English-speaking countries" which know best how English should be taught. (Romaine 2006: 596)[6]

However, the gatekeeping function of an exonormative standard is not only apparent in the political and cultural relations between the former colonizers and the former colonized; it is also called upon by local elites within the Outer Circle speech communities:

> the perceived linguistic differences ("native" and "non-native") become indices of social positions of speakers and reflections of the quantities of linguistic, cultural, and/or symbolic capital they possess. The more linguistic capital speakers possess, the more they are able to exploit the system of differences to their advantage (native-standard/nonnative-nonstandard) and thereby secure a profit of distinction. (Bhatt 2007: 57)

In that sense, the English-speaking elites in the Outer Circle capitalize on the "standard ideology" in Milroy's sense, which separates the "legitimate" standard variety from all others (Milroy 2001: 547). The irony here is that adherence to the standard ideology will always marginalize Outer Circle speakers of English, however hard they try: the division between "the legitimate and illegitimate offspring of English" (Mufwene 2001: 106–125) rests on the concept of "native speaker" in Quirk's sense, so that speakers in the Outer Circle are by definition disenfranchised.

On one level of description, then, the equation "native = standard = higher stakes in the linguistic marketplace" surely holds, and the debates about the role of and access to English in a globalizing economy are likely to continue. For our purposes, it is also necessary to narrow down the perspective and to focus on the ideology within linguistic theory with respect to "the native speaker".

6. The implications of the native/non-native distinction for the worldwide teaching of English are another widely discussed topic, which cannot be addressed here.

2.2.2 The native speaker and linguistic variation

Recall from the preceding section that Quirk (1990) drew a categorical dividing line between native and non-native varieties of English. Among the former, he recognized a further distinction between institutionalized (i.e. codified) varieties such as British and American English and non-institutionalized varieties such as e.g. Australian and New Zealand English (1990: 6). The latter varieties were not amenable to such a further distinction: as learner varieties, they are by definition outside the scope of institutionalization, since they represent only intermediate stages on the way to full proficiency in the target language, that is BrE or AmE.[7] Quirk's position captures the issue involved in conceptualizing IndE and other New Englishes as varieties:

> No one should underestimate the problem of teaching English in such countries as India and Nigeria, where the English of the teachers themselves inevitably bears the stamp of locally acquired deviation from the standard language ("You are knowing my father, isn't it?") The temptation is great to accept the situation and even to justify it in euphemistically sociolinguistic terms. [...] It is neither liberal nor liberating to permit learners to settle for lower standards than the best, and it is a travesty of liberalism to tolerate low standards which will lock the least fortunate into the least rewarding careers. (Quirk 1990: 8f.)

Quirk's rather passionate stance here is that of the applied linguist and dedicated language teacher, and this approach is also apparent in theoretical linguistics, as Singh has noted:

> the very notion of non-native phenomena in English is predicated on the legitimacy of the error-driven, pedagogical enterprise that studies what it calls non-native varieties of English. (2007: 34)

Quirk's phrase "locally acquired deviation from the standard language" may be taken to illustrate some of the pitfalls that loom large in the study of New Englishes, pitfalls which shall be avoided in my study of IndE. First of all, the New Englishes have been conceptualized within the framework of second language acquisition (SLA), and while it is undoubtedly true that New Englishes more or less by definition are second languages, the "error-driven, pedagogical enterprise" of SLA has imposed a perspective on variation within the New Englishes that is predicated on the existence of an exonormative target language. The division into "native" and "non-native" implies that variation in the former's linguistic system is an indicator of a separate subsystem or of emerging dialect change.

7. This probably prompted Kachru to insist on the term "institutionalized non-native variety of English" for IndE, cf. e.g. Kachru 1986: 92.

Variation in the latter speech community may be either simply performance errors in the process of language acquisition or fossilized interlanguage features, in any way indicators of "arrested development" or imperfect language learning rather than systematic variation. In other words, scholars might be predisposed to consider a given realisation of a linguistic variable as an innovation worthy of attention if they are dealing with "Old" Englishes, but as an idiosyncratic mistake (and as such to be neglected) if it occurs in New Englishes. The SLA perspective is also somewhat myopic in focussing mainly on the performance of the individual speaker/learner rather than on the speech community to which he/she belongs, and from which variation derives its social significance. If all speakers of English as a second language are conceptualized as learners, this narrow perspective is a natural consequence of the SLA framework. Within that framework, "locally acquired deviation" may take one of two forms: it may represent an intermediate stage in the learner's acquisition process, soon to be overcome when the learner's efforts lead him/her closer to the target language. If the "deviation" persists, then the learner has failed to reach the target and the "deviation" becomes "fossilized", part of his/her interlanguage. Sridhar argued early on against the conceptual limitations of applying the SLA approach to the New Englishes in their multilingual settings:

> The occasional references to the second language contexts have implied that varieties such as Indian English are fossilized stages in the acquisition of native speaker norms [...]. They also suggest that mother tongue influence is a characteristic of early stages in the acquisition process that is eventually overcome by the ideal learner [...]. These assumptions result from an inadequate appreciation of the sociolinguistic context in which SLVs [second language varieties] are acquired. As we have seen, the transference of patterns (lexical, pragmatic) from the mother tongue serves the important function of adapting an alien code to the socio-cultural context of use. Secondly, as several scholars have pointed out, a speaker who uses native-like pronunciation (as well as other features such as hyper-correct grammar) is viewed with suspicion and derision in second language contexts. (1985: 52)

What is an impediment to successful language learning from the SLA perspective, namely interference from the mother tongue, becomes an asset in this view: the multilingual speaker can freely draw on the resources of his/her overall linguistic repertoire in order to "nativize" the second language, to use Kachru's familiar term. Sridhar also explicitly refers to 'transfer of pragmatic patterns'; this point will be taken up again in Chapter 3.2.4 on South Asia as a sociolinguistic area.

Mesthrie has further challenged the notion of fossilization as too static to capture speakers' competence in exploiting different varieties of English along the cline of proficiency for stylistic effects:

> The problem arises when one fails to draw a distinction between process (of language learning) and state (a new, but relatively stable, dialect). For many New Englishes speakers backsliding becomes a stylistic feature, rather than the re-emergence of an error. It is part of the competence of the speaker that enables him/her to choose between variables, rather than an aspect of (imperfect) performance. (1992: 154)

These authors – among many others – advocate abandoning the "deviationist approach" (Sridhar 1985: 44), which entails an unwarranted reliance on SLA conceptualizations, and assert that the notion of "deviation" with reference to New Englishes has to be replaced by "variation": "The relevant question is whether this variation is patterned and can be related to the customary social and linguistic parameters" (1985: 47). This question can only productively be asked if speakers of New Englishes are no longer seen as learners, but as members of a speech community. Labov's much-quoted definition of "speech community" runs as follows:

> The speech community is not defined by any marked agreement in the use of language elements, so much as by participation in a set of shared norms; these norms may be observed in overt types of evaluative behavior, and by the uniformity of abstract patterns of variation which are invariant in respect to particular levels of usage. (Labov 1972a: 120f.)

The crucial point here is that members of a speech community are bound by *awareness* of a set of shared norms: the community's shared norms provide the reference point for the evaluation of the social meaning of "orderly heterogeneity", to name another classic sociolinguistic concept. The ensuing crucial question then brings us back to the 'native speaker': does it take native speakers – and only those – to form a speech community with shared norms? Does the Indian English speech community consist of the whole population, the estimated 350 million speakers who use English on a regular basis, a third of the population (cf. Crystal 2008: 5), or the 226,449 people (0.02% of the population) who returned English as their mother tongue in the 2001 census (cf. Chapter 4.1.3.3)? Few scholars writing in this context would join Paikeday (1985) in his triumphant chorus "The native speaker is dead!" and abandon the notion altogether; most would probably follow Singh in his assessment: "Although it is instructive to deconstruct certain construals of *native speaker/user*, not much is to be gained by discarding the concept completely" (2006: 489). What needs to be deconstructed, then, is the dichotomy "first language – second language", where only the "first language" or "mother tongue" confers "native" competence upon the speaker. This dichotomy incorporates a monolingual bias which is inadequate for most of the world's speech communities, and it cannot be sustained for speech communities where grassroots multilingualism is the norm rather than the exception: an Indian child may well grow up to

speak three to four different "mother tongues" even before formal schooling (cf. Pattanayak 1998: 128).

Terms such as "first language" and "second language" imply successive rather than simultaneous acquisition, such that only one language becomes the "native" language, the language which marks the boundaries of the speaker's speech community. Again, such a view is highly ideological in epistemologically privileging categories predicated upon monolingualism as the "normal" state of affairs, and multilingualism as "deviant". Chapter 3 will attempt to chart the Indian communicative space and its patterns of multilingualism, which are not likely to yield to monolingualism any time soon. What is needed, then, is the recognition that a "bilingual speaker is NOT a simple, additive union of two monolingual speakers" (Singh 2007: 37): the "functionally determined distribution of the use of particular languages and the concomitant acquisition and competence in them in multilingual societies" (2007: 36) renders the monolingual perspective inadequate. Research on bilingual speech communities indeed lends support to the view that a bilingual speaker's repertoire is more than the sum of its component parts, in this case the speaker's individual languages. Ever since Poplack's classic study on a bilingual Puerto-Rican speech community in the US (Poplack 1980), evidence has been accumulating that bilingual speech patterns such as code-switching constitute a separate discourse mode with its own norms and distinct from the norms of either of the two individual languages (cf. also Gumperz & Cook-Gumperz 2005).

Singh then suggests "a characterization grounded squarely in the reality and psycholinguistics of multilingualism" (2007: 39) for the notion of *native speaker/user*:

> A native speaker/user is, in other words, a speaker/user whose well-formedness judgements on utterances said to be from her language are shared by the community she can be said to be a member of. (2007: 38)

To be sure, adopting a multilingual perspective on the first/second language dichotomy disposes of some problems, but not all, as Singh himself admits when he offers his definition of 'native speaker/user': "The only thing to remember is that we are talking about speakers and NOT learners" (2007: 43), which immediately raises the question whether we have principled, *a priori* criteria for distinguishing between the two. Mesthrie's study of language shift in a community of South African Indians (1992) may serve as an example of how to approach this problem: he assigned the speakers taking part in his survey to the familiar groups "basilectal – mesolectal – acrolectal speaker" first and foremost on the basis of his intuition as a native speaker of the variety in question. Further, the dominant factor for assigning individual speakers to one of the three groups is the speaker's level of education. Thus, Mesthrie's own membership in the speech

community in question enabled him to make judgements about the proficiency level of other members "according to the established norm of the community" (Mufwene 1998: 117).

We see now that to discard of the notion "native speaker" altogether is not helpful; we may still need his/her intuitions that are based on his/her access to the speech communities' norms. It is important here to carefully disentangle the two related, but distinct notions of "norm" and "standard", as Koch (1988) has convincingly demonstrated. All speech communities develop norms, but not all speech communities develop one of the many available descriptive norms to the point where it becomes the sole prescriptive norm, i.e. the standard. When Singh points out that IndE should be studied on a par with other non-standard varieties of English, he focuses on this neglected conceptual difference between norms and the standard, where "standard" is implicitly equated with (highly) educated and mostly idealized usage:

> I AM painfully aware of the fact that being a native speaker is no protection against discrimination, but I draw some consolation from the fact that my modest demonstration that speakers (=/ learners) of Indian English are native speakers of Indian English just as speakers of Midwestern American English are native speakers of Midwestern American English not only establishes what needs to be established but also makes it clear that Indian English needs to be studied with all the opposites of standard British and American English. (Singh 2007: 43)

It is probably more accurate to say that IndE is a *pre*-standard rather than a *non*-standard variety; unlike Midwestern American English, whose norms will not be developed into a standard because the general AmE standard is already in place, IndE may well proceed to a stage where a particular norm and the standard become co-extensive, as has already been argued in the preceding section (cf. also 2.4 below).

This leads us back to the quote from Quirk concerning "deviation from the standard language". Many studies have unquestioningly taken written Standard British English as the basis for comparison rather than actual language use. As such, an idealized, monolithic and static "standard English", its rules derived by introspection, was compared with a performance variety. Earlier studies on IndE typically presented lists of "errors" or "deviations" to "native speakers of English" (i.e. speakers of British or American English) to elicit their acceptability judgements concerning IndE usage (e.g. Parasher 1983). More often than not, these "deviations" were also prefabricated examples of "typical Indianisms" rather than naturally occurring data. Mehrotra's study "A British response to some Indian English usages" (2003) is a rather late example of this approach, as the title already indicates: even though Mehrotra's overall intention is to question the authority of

the "native speaker" in Quirk's sense, the very design of his study moves the "native speaker" centre stage as the one who passes judgment on intelligibility. Another group of studies, no doubt inspired by Kachru's much-criticized precedent (cf. the preceding section), took its examples of "Indianisms" from Indian novels written in English (e.g. d'Souza 1991b, Pandharipande 1992).

Studies such as the one by Sharma (2001) then proved a welcome addition to a field which was in dire need of a more empirical orientation: Sharma made a strong case both for focussing on actual language data and for acknowledging the inherent variability in written American and British English alike. In her study, she compared "quantitative and qualitative aspects of pluperfect use" (2001: 343) in three parallel corpora of written English, namely the Brown corpus (AmE), the LOB corpus (BrE) and the Kolhapur Corpus (IndE). As one of her results, she highlighted "the importance of taking into account variability and ambiguity in the lexifier or native variety itself when considering possible sources of new form–function relationships in a new variety" (2001: 366).

As more and more corpora within the general format of the ICE-project become available (cf. Section 4.1.2), unbiased comparative studies can now be extended to spoken language as well, and the differences between text types and genres can be taken into account. The ICE-project, then, has certainly paved the way for a more empirical approach to the study of New Englishes as varieties in their own right.[8]

2.2.3 Outlook

This section has only touched upon some aspects involved in the distinction "native" vs. "non-native" and the corresponding distinction "first language vs. second language". The question if – and how – these distinctions should be conceptualized in a theoretical framework will continue to be highly relevant for the World Englishes community. Most researchers would probably agree that it was a necessary undertaking to deconstruct the ideological underpinnings of the 'native speaker'-concept. As I see it, there is no need to discard of the concept altogether; the important theoretical advance resulting from the debate comes from severing the link between 'native speaker' and monolingualism, which was largely taken as an axiom of linguistic thought.

Broadly speaking, current research moves in two different directions: scholars such as Mufwene (2001, 2008) and Schneider (2007b) propose to analyse postcolonial varieties of English including Pidgins and Creoles within a unified

8. Of course, I do not want to imply (a) that all corpus-based studies are by definition unbiased, and (b) that only corpus-based studies deserve the label 'empirical'.

framework in which the original settlement patterns and contact scenarios play a decisive role. For Schneider, "recent realities seem to be rendering the ENL-ESL distinction increasingly obsolete" (2007b: 13); that is the distinction between English as L1 or L2 is becoming increasingly blurred, the faultline is between ENL/ESL and EFL. Mesthrie, albeit jokingly, even goes one step further and comments on "the blurring of the circles in the twenty-first century":

> Everyone knows – or ought to know – that, give or take a few exceptions, the real difference between Outer Circle and Expanding Circle is whether you play cricket or not. (2008a: 32)

The editors of the *Handbook of Varieties of English,* on the other hand, maintain a threefold distinction between L1 Englishes, L2 Englishes, and Pidgins and Creoles, and Kortmann and Szmrecsanyi (2004) offer evidence that there is a distinction to be drawn between L1 and L2 Englishes on the basis of specific structural features/properties. According to their survey, differences between the "Old" and the "New" Englishes are not simply quantitative, as Schneider (2007b) and others argue, but also qualitative: some features occur exclusively in L1s and L2s, respectively. It has to be kept in mind, though, that Kortmann and Szmrecsanyi use a rather broad brush for painting their picture, as Sharma (2009) has recently demonstrated: her in-depth study of some of those features in Asian Englishes which rank high on Kortmann and Szmrecsanyi's list of "candidates for universals of New Englishes" (2004: 1193) supports the following conclusion:

> Properties that may appear to be "typical of L2 varieties" may in fact be parallel cases of transfer of typologically common features in the substrates, or may be differently conditioned in each variety. (Sharma 2009: 173)

Thus, "highly abstract aggregations of varieties in the search for universals" (*ibid*. 192) may obscure rather than illuminate processes of contact-induced change and unwittingly reify the distinction between English as first and as second language. The new *Electronic World Atlas of Varieties of English* (eWAVE), intended as a follow-up rather than a companion website to the *Handbook of Varieties of English*, has greatly increased the number of features covered and has further introduced a more fine-grained distinction for the classification of varieties, but only where L1 varieties are concerned: "Old" Englishes can now be categorized as either "high-contact L1" (e.g. Irish English, Colloquial Singapore English) or "low-contact L1" (e.g. Scottish English, Appalachian English) (http://www.ewave-atlas.org/welcome/introduction). It remains to be seen in how far these amendments may serve to move beyond the kind of sweeping generalizations that were criticized by Sharma (2009). Research strategies which try to avoid some of the conceptual traps outlined above will be the topic of the next section.

2.3 Contact-induced language change

2.3.1 Introduction

The two preceding sections have sought to trace how some core concepts of the field have evolved over the years – concepts which, although fundamental to the very pursuit of research on New Englishes, never went uncontested. Some of the conceptual and ideological challenges that scholars of New Englishes directed at their colleagues have had interesting repercussions in other fields, for example historical linguistics: it has been argued that *all* varieties of English in all periods have been contact languages (Mesthrie 2006a, Mufwene 2008) and should be studied as such. Thomason has made a similar point: for her, historical linguistics deals with dialect borrowing or feature diffusion within speech communities, and contact linguistics focuses on feature diffusion across language boundaries, "but the differences turn out to be a matter of degree, not of kind" (2003: 688). Eventually, this means that some of the traditional explanatory parameters for linguistic change in the history of English have to be recast, namely the notions of "internal" vs. "external" motivation for language change. "External motivation" or language contact should no longer be a last resort for explaining a specific phenomenon, to be evoked when all else fails, but should be acknowledged without monolingual bias in the conceptualization of language change.

As I have indicated at the beginning of this chapter, New Englishes are contact languages *par excellence* and offer outstanding opportunities to put current theories on the mechanisms of contact-induced language change to the test. Curiously, and perhaps inevitably, given the history of the field, most of the research on the New Englishes over the last decades more or less consciously deprived itself of a larger perspective. Rather than treating New Englishes as contact varieties among many others, scholars tended to emphasize their unique status, even going as far as postulating some unique explanatory parameters such as "angloversals" (cf. Mair 2003) for those features the New Englishes allegedly have in common and which supposedly distinguish them as a group from the "Old" Englishes. When d'Souza in 1992 implicitly criticized the emphasis on the uniqueness of the New Englishes as an object of study, her comments seemed to have been overlooked: she argued that the prevalence of code-mixing in the New Englishes does not constitute a defining characteristic, but is "an inevitable outcome of any bilingual context and, therefore, only incidental to the description of an NVE [new variety of English] *per se*" (d'Souza 1992: 217). She concludes:

> In dismissing the claim that CM [code-mixing] is a defining characteristic of the NVEs, it may seem that I am stripping the NVEs of one of their most distinctive features. This is not at all the case. My claim is that CM is the result of bilingualism and language contact, and will be found in all such contexts even when English does not play a part in that context. It is an accident of history that English has become a world language and one result of its spread is that it plays an important part in mixing. Mixed varieties arise even in non-NVE contexts. To tie CM to the study of the NVEs is to blind ourselves to many crucial aspects of both phenomena and to claim that the NVEs are in some essential way different from other codes. (1992: 222)

With this insightful conclusion, d'Souza has prefigured the 21st-century trend towards discarding the categorical distinctions between "New" and "Old" Englishes as well as the impetus towards developing a unified framework for the characterization of World Englishes, prominent among them Schneider's Dynamic Model of Postcolonial Englishes (cf. Chapter 2.1.4). When she singles out code-mixing as one of the most conspicuous features of New Englishes, she addresses one of the topics that have preoccupied all scholars working on language contact, namely the exact status of code-mixing (*vis-à-vis* borrowing) and its role in contact-induced language change. In what follows, this question will feature prominently.

There are already more theoretical perspectives on language contact than can be given proper credit here: the models put forward by Thomason & Kaufman (1988) and later Thomason (2001), van Coetsem (2000), Myers-Scotton (2002), Johanson (2002), Heine & Kuteva (2005), Matras (1998, 2000) and further Matras & Sakel (2007) have all significantly advanced our understanding of language contact.[9] My main concern in the following will be to focus on some of the theoretical and methodological assumptions derived from contact linguistics that inform the present study. Two important caveats have to be kept in mind, however: first of all, "great intensity of contact is a *necessary* condition for certain kinds of interference, especially structural interference, but it is by no means a *sufficient* condition" (Thomason 2003: 688). In much of current research, Weinreich's notion of "interference" has in a way been reinstated as the default mode of multilingual communication:

> Transfer and interference have traditionally been regarded as negative elements in the learning process and as manifestations of failure to acquire the correct structures of the target language. But as long as they do not result in incomprehensibility and a breakdown of communication, one might instead view them as *enabling* factors that allow language users to create bridges among different subsets within their overall repertoire of linguistic forms, and to use these bridges to sustain communication. (Matras 2009: 74)

9. Backus 2005, Winford 2005, Matras 2009 provide a comparative overview of selected approaches.

Thus, contact-induced language change in multilingual settings now appears less in need of explanation than its absence. However, "stability of structure, i.e. demonstrating what does NOT change, is an overlooked but important topic for any theory of change" (Backus 2004: 180). Thomason further points to

> the idiosyncratic nature of speakers' creativity. This factor alone guarantees the continuing failure of all attempts to construct a neat hierarchical ordering valid for all types of contact-induced change. (2003: 709).

This point, however, is valid for all theories directed at explaining language change, be they contact-induced or not. In Lass' terms (1980: 131):

> I would claim then that linguistic change is entirely a domain of options, including the zero options. *No change is ever necessary.* If it were, it would already have happened everywhere.

Another point raised by Thomason impinges more specifically on the present study: models of contact-induced language change are typically predicated on contact scenarios involving only two languages; the complexities of patterns of stable multilingualism and of longstanding *sprachbund* situations may well be beyond the current models (cf. Thomason 2001: 99). Thomason herself has already refined the original model put forward by Thomason & Kaufman (1991 [1988]), which distinguished two major types of linguistic interference in language contact situations. According to Thomason & Kaufman, "borrowing" occurs under conditions of full bilingualism and typically leads to the borrowing of non-basic vocabulary. Structural features are borrowed, if at all, if the contact between the two speech communities is sufficiently intense (cf. Thomason 2003: 691). "Shift-induced interference", on the other hand, first targets phonological and syntactic features and typically happens in the process of imperfect learning of the target language (TL) by the shifting speech community.

Anybody familiar with the literature on the New Englishes is likely to raise two objections here: firstly, the notion of "imperfect learning" is highly reminiscent of the more traditional SLA-approach to the New Englishes discussed in Section 2.2.2, which characterized New Englishes as fossilized stages of an imperfectly learned target language. Secondly, and perhaps more importantly, most New Englishes flourish under conditions of stable multilingualism, and structural interference occurs without any indications of language shift. Thomason has acknowledged these objections, with specific reference to the Indian situation, and tried to incorporate them into the original model:

> It is important to keep in mind that imperfect learning in this context does not mean inability to learn, or even lack of sufficient access to the TL to permit full learning: learners must surely decide sometimes, consciously or unconsciously, to use features that are not used by native speakers of the TL.

> Another point that must be made emphatically is that this type of interference can occur without language shift. In India, for instance, there is a variety of English known as 'Indian English' that has numerous interference features of this type from indigenous languages of India; Indian English is spoken by many educated Indians who speak other languages natively. So, although it is a variety that is characteristic of one country it is not, strictly speaking, a variety formed under shift conditions. The references [...] to language shift and shift-induced interference therefore carry an implicit warning label: the linguistic predictions are the same for all instances of imperfect group learning of a TL regardless of whether or not actual shift has occurred. (Thomason 2001: 74)

Thomason proposes to retain the term "shift-induced interference" simply for want of a better term, even though language shift is neither a necessary nor a sufficient condition for structural interference (Thomason 2003: 692). It remains to be seen whether more complex multilingual contact scenarios such as the Indian case will necessitate further revisions in existing models of contact-induced language change.[10]

2.3.2 Mechanisms of contact-induced language change

Most researchers on language contact agree that the individual multilingual speaker in actual contact settings should be given pride of place in any further consideration of mechanisms of contact-induced language change. The logic of focussing on what speakers actually *do*, whether multilingual or not, was already expounded in Weinreich et al.'s classic article on language change in general (1968). The authors also elaborated on the fundamental problem of studying the actuation of language change, namely that "all explanations [...] will be after the fact" (1968: 186). Their famous dictum illustrates the problem: "Not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity" (1986: 188).

Thomason's model of language contact and contact-induced language change has tried to reconcile the emphasis on the individual speakers' behaviour with a broader historical perspective on language contact scenarios. The latter informs the well-known dichotomy of borrowing vs. shift-induced interference that has already been alluded to above. The most influential aspect of this model is the

10. Masica's insightful discussion of constraints on borrowing with respect to the South Asian evidence even calls the empirical foundation of the basic distinction between borrowing and shift-induced interference into question: "there was already important external influence on early Indo-Aryan [...] without much lexical borrowing" (2001: 231). He concludes that the complexity of linguistic areas such as South Asia, where more than two languages are in contact over millennia, may well elude a model which relies on a plain binary distinction (2001: 234).

Table 2.2. Thomason & Kaufman's borrowing scale (adapted from Thomason 2001: 70–71, quotes *ibid*.)

Degree of contact	Effect on the lexicon	Effect on structure
casual contact	content words (nonbasic vocabulary)	none
Slightly more intense contact	content words (nonbasic vocabulary) as well as function words	minor structural borrowing: phonological features in loanwords, extension of functions for existing syntactic structures
More intense contact	more function words, basic vocabulary, derivational affixes	moderate structural borrowing: phonology, morphology, word order, syntax of clause-combining
Intense contact	heavy lexical borrowing	"Anything goes, including structural borrowing that results in major typological changes in the borrowing language"

equally well-known borrowing scale (Thomason & Kaufman 1988: 74–6, Thomason 2001: 20f.) depicted in Table 2.2.

The claim that "anything goes" in language convergence, given the appropriate social conditions, has been the most controversial issue within Thomason & Kaufman's overall framework. The attempt to identify constraints on contact-induced language change has been among researchers' prime concerns, and Thomason's insistence that the main determinant for the outcome of language contact is the social setting and that there are no absolute constraints has been heavily criticised.[11]

Her main argument against the possibility of finding absolute constraints on contact-induced language change is worth extensive quoting:

> A linguistic innovation may or may not have a strictly linguistic motivation: it may arise entirely through processes inside the head of a monolingual or bilingual person, in which case it arguably has only a linguistic trigger, or it may arise from something that a monolingual or bilingual person hears from another speaker (e.g. a misperception of a particular speech sound or word), in which case it results in part from social interaction. Even if the innovation is entirely linguistic in nature, however, the spread of any innovation through a speech community must certainly be social at least in part, because it is governed by such things as social networks, prestige, and various demographic factors. The spread of a structural innovation is also very likely to depend in part on linguistic factors; the likelihood of its successful spread is enhanced if it is less marked (easier to learn), if it fits well typologically with the language's pre-existing structural patterns, and if it fits neatly into the language's internal tendencies of drift having to do with inherited structural imbalances. Nevertheless, the spread of an innovation cannot

11. Thomason (2008) provides an overview of the discussion.

> be due entirely to linguistic factors. *Any search for absolute linguistic constraints on language change, whether internally-motivated change or contact-induced change, must focus on the process of innovation by one or more individuals.* Even the most powerful linguistic tendencies that favor or disfavor the spread of a given innovation cannot be absolute constraints; if they were, the innovation (ephemeral or permanent) could not have occurred in the first place. (Thomason 2008: 50–51, emphasis mine)

This brings us back to the agentivity of individual speakers in multilingual settings[12] – an aspect of language contact that is included in Thomason's theory, but not as elaborated as e.g. the borrowing scale.

Thomason lists seven mechanisms of contact-induced language change (Thomason 2001: 129–156), among them e.g. code-switching, code-alternation, bilingual first language acquisition, and second-language acquisition strategies. All of these mechanisms are predicated on speakers' strategies in multilingual communicative situations. Although other researchers have proposed different sets of mechanisms (e.g. Matras 2000b, Johanson 2002, Backus 2005, Heine & Kuteva 2005), practically everyone working in the field acknowledges the importance of code-switching as a possible mechanism of contact-induced language change. Backus (2005), for example, considers code-switching as a process of language change. For him, code-switching and structural borrowing are intimately related: "CS [code-switching] and structural borrowing must be instantiations of the same process" (Backus 2005: 315). The literature on code-switching "as a discourse mode of a bilingual community" (Poplack & Meechan 1998: 128) is vast, and some of its preoccupations are similar to those that drive research on contact-induced language change.[13] Two issues are particularly relevant in our context. Firstly, the issue of constraints on code-switching has been discussed extensively. Such synchronic constraints on possible surface realizations of bilingual behaviour will also constrain possible long-term changes in a contact scenario. I will come back to this point below in the discussion of Matras & Sakel's proposal (2007) for a primarily cognitive model of contact-induced language change. Next, a "discussion that will not go away concerns the issue of distinguishing borrowing from code switching", as Myers-Scotton (2002: 153) wryly noted. For her as for many others (e.g. Thomason 2001: 133–136), the distinction between borrowing and code-switching is rather fuzzy and one of degree rather than of kind. Further, the distinction can only be made in retrospect; what starts life as a code-switch in

12. The terms RL [recipient language] and SL [source language] agentivity play a major role in van Coetsem's framework (2000).

13. For a comprehensive overview of the state of the art in code-switching research see Gardner-Chloros 2009.

the interaction of fluent bilinguals may turn into a borrowed item once it has become established in the overall speech community who does not necessarily know the source language. A borrowing is then just a code-switch that has come of age, so to say. However, the distinction is categorically upheld in the work of Poplack and her associates (e.g. Poplack 1980, Poplack & Sankoff 1984, Poplack & Meechan 1998), where the constraints on code-switching identified by Poplack (1980) are said to hold for 'true' code-switches only and not for borrowings. This topic will be taken up again in the context of Chapter 4.6.2. For the time being, Matras' criticism of the term 'borrowing' indicates a different perspective on the issue that moves away from static categorizations (2009: 146):

> One might thus contend that 'borrowing' emphasises too much the aspect of ownership and the boundaries between the linguistic systems involved, and that this diverts attention away from the dynamic process of sharing a structure or word-form, adopting, applying, and using it. Johanson (2002) therefore favours the term 'copying', which emphasises the creative use of an item within the 'recipient' language. I will use the term *replication* to capture even more closely the fact that we are dealing not with issues of ownership or even direct imitation or duplication, but rather with the activity of employing an item, in context, in order to achieve a communicative goal.

The notion of "replication" in the sense of "*replication* of a linguistic structure of any kind, in a new, extended set of contexts, understood to be negotiated in a different language" (Matras 2009: 146) figures prominently in the model of contact-induced language change put forward by Matras & Sakel (2007). In the authors' terminology, "replication" covers two distinct types: 'matter replication' (or MAT for short) is defined as "direct replication of morphemes and phonological shapes from a source language" (Matras & Sakel 2007: 829). Pattern replication or PAT refers to the

> re-shaping of language-internal structures. In the latter process, the formal substance or matter is not imported but is taken from the inherited stock of forms of the recipient or replica language (i.e. the language that is undergoing change). Rather, it is the patterns of distribution, of grammatical and semantic meaning, and of formal-syntactic arrangement at various levels (discourse, clause, phrase, or word) that are modelled on an external source. (2007: 829f.)

This distinction has a long tradition in the study of language contact (cf. Matras 2009: 236) which the authors readily acknowledge. The distinction is relevant because when "it comes to matter units, speakers seem to be on the whole more conscious of their selection" (Matras 2009: 148), that is, speakers and speech communities are much more aware of imported matter than of imported patterns, a factor that might influence the trajectory of contact-induced language change. Matter and pattern

replication may also target different linguistic levels (cf. Matras & Sakel 2007: 844–847), another point that will be taken up in subsequent chapters.

Again in line with previous work in the field, the authors focus on the actual multilingual communicative situation as the arena where language contact is played out in speakers' face-to-face interaction. What is new in their approach is that they locate the trigger for contact-induced language change in speakers' cognitive strategies:

> In order to understand the triggers behind various mechanisms of change in situations of contact, we must therefore explore multilingual speakers' motivations to adopt ad hoc solutions and strategies in response to the communicative challenges that face them when structuring individual utterances in discourse. Our agenda is therefore to trace the seeds of pattern replication; we propose to search for them in the motivations which speakers display for adopting utterance-level solutions to the communicative challenges of multilingual settings. (Matras & Sakel 2007: 848)

Such a conceptualization of replication as the result of bilingual speakers' cognitive processing operations owes much to Matras' earlier work on the borrowing of utterance modifiers, a category that will be extensively discussed in Chapter 4.6 below. Matras (1998, 2000a, 2000b) noted that utterance modifiers, a "somewhat extended group of operators that are responsible for monitoring and directing the hearer's processing of propositional content" (2009: 137) proved particularly prone to replication across a range of diverse language contact scenarios. Among the prototypical members of this class are discourse markers and focus particles as well as conjunctions, tags, and interjections. Matras observed that such items are typically derived from the "pragmatically dominant language" (2000b: 577):

> the language which, in a given moment of discourse interaction, is granted maximum mental effort by speakers. This may be the speaker's first language, or one that is dominant for a particular domain of linguistic interaction, or one that exerts pressure due to its overall role as the majority language that is culturally prestigious or economically powerful. (*ibid.*)

The notion of "pragmatically dominant language" is flexible and context-dependent and thus better suited to multilingual settings, where the pragmatically dominant language might change from one conversational encounter to the next. Unlike Thomason's concept of a dominant language as the target language in shift-induced interference, this concept does not postulate language shift as a precondition for convergence. The empirical problem that Thomason acknowledged for her own model and discussed above, namely that structural interference in Indian English is observable without language shift, would not arise with this concept.

Utterance modifiers derived from a pragmatically dominant language generally enter a target language via matter replication, and there will be ample opportunity

to discuss evidence for this process in spoken IndE in Chapter 4.6. Pattern replication, on the other hand, crucially involves what Matras & Sakel (2007) call "pivot-matching". Multilingual speakers may "give up the separation between their "languages" – the mental demarcation line that divides their overall repertoire of linguistic structures – in respect of a particular function-bearing structure (a "category")" (Matras 2007b: 31). The motivation for doing so is cognitive:

> The advantage to the speaker [...] is the syncretisation of processing operations in the two languages, allowing speakers to apply similar mental organisation procedures to propositions in both languages of their repertoire. (Matras & Sakel 2007: 835)

The process of pivot-matching as Matras & Sakel conceptualize it can be illustrated with one of their examples. A German tourist with only little knowledge of English wants to express that a specific place was closed:

(2.1) a. Tourist's English:
It was to.

b. German:
Es war zu.
it was to
'It was closed.' (Matras & Sakel 2007: 851)

"Rather than refrain from communication, the speaker makes use of the full range of constructions available in her repertoire" (*ibid.*). She does not seem to know the English word 'closed', but still wants to express herself in English and then draws on the resources of her mother tongue while at the same time observing the constraint on the current communicative situation, namely that the language of interaction is English. According to Matras & Sakel, the speaker identifies *zu* as the pivotal feature of the German construction and tries to replicate what she perceives as the matching pivot in the target language English. German *zu* is highly polysemous. The speaker has exploited the polysemy of *zu*, including the adjectival and the prepositional meaning, and chosen the English preposition *to* as the closest equivalent in the replica language that was available to her. Phonological similarity may also have played a role here, even though Matras & Sakel consider this marginal for the process of pivot-matching (2007: 856). According to them,

> The most powerful force in 'pivot-matching' appears to us to be the semantic potential of a structure in the replica language to cover the (lexical or grammatical) semantics represented by the model. The inspiration or clue which points the speaker in an appropriate direction is the polysemy or polyfunctionality of the model. This search for potential clues leads the speaker quite naturally and frequently to more concrete, semantically 'basic' structural elements within the replica language, rather than to more abstract ones. (2007: 852)

The fact that the German tourist's utterance is downright wrong by any standard is irrelevant: at the level of actuation, all novel utterances and structures that speakers produce in multilingal settings are created equal.

Matras & Sakel clearly "distinguish between language convergence, as an innovation triggered at the level of communication between a speaker and a hearer, and the propagation of innovations leading to language change" (2007: 849). In that sense, their argument closely parallels the point made by Thomason as quoted above. Innovation happens in multilingual face-to-face interaction and is driven by

> the need to perform effectively in communicative interaction while adhering, on the one hand, to the rules about the selection of clearly-identifiable phonological substance (matter) from the language that is appropriate in the particular context, while at the same time exploiting constructions that are available to the speaker in his/her entire repertoire of linguistic-communicative structures. (Matras & Sakel 2007: 854)

The processes that are active here are the same for fluent bilinguals, second language learners with limited language skills, or children acquiring more than one language. The actual innovations produced, however, do not stand the same chance of propagation, which is fundamentally a social process of negotiation within and across the speech community. Mistakes such as *it was to* in the German tourist's utterance are unlikely to stabilise, but

> it is precisely the sociolinguistic limitations, and not the structural character of the replica construction, or the fashion in which it emerged (spontaneously), that are the crucial factors which prevent propagation and hence change. (2007: 851)

Thus, the model allows to dispense with the notorious problem of distinguishing "mistakes" and "innovations" in the New Englishes, and ultimately also with the equally problematic distinction between native speakers' and non-native speakers' competence. At the performance level, all phenomena are motivated by the same cognitive pressure on speakers in multilingual settings. Differentiation into "mistakes" vs. "innovations" only becomes relevant at the level of propagation, not at the level of actuation. Interestingly, a similar argument has already been made in passing by S. N. Sridhar with specific reference to the Indian context (1992: 149):

> by extending the rules of the mother tongue to the other tongue, the speaker minimizes cognitive dissonance and maximizes cognitive economy. And because the interlocutors are also bilingual and share the same rule systems, such transfer does not interfere with intelligibility. As I have stated elsewhere, transfer is the grease that makes the wheels of bilingual communication run smoothly. One might go

> further and even suggest that such transfer, by polishing off the rough edges of the foreign language rule system (such as the use of the expletive *there*, the variable tag in yes-no questions, and so on, from the point of view of the Indian speaker), and reducing the differences, makes the use of the other tongue more acceptable in a bilingual situation.

Sridhar does not only allude to the actuation of convergence features, but also to the likelihood of their stabilisation. Interference or transfer may serve "the important function of adapting an alien code to the socio-cultural context of use" (K. Sridhar 1985: 52), that is, the variety's acceptability may actually be enhanced by incorporating such interference features. The next chapter will consider the relation between norms and standards for English in multilingual speech communities in more detail. The other noteworthy feature of Matras & Sakel's model is the characterisation of the process of convergence via pivot-matching, which allows more precise descriptions of the outcomes of language contact than before. More detailed discussions will follow in Chapters 4.2 and 4.3.

2.4 Norms and standards

The preceding Chapters 2.1 and 2.2 have devoted much space to the changing status of New Englishes in general and IndE in particular – from "deviant" or "fossilized" varieties to varieties of English in their own right. However, this change in attitude, which marks the endpoint of long theoretical debates, seems to be restricted to the experts, i.e. the linguists interested in World Englishes. A recent undergraduate textbook on IndE illustrates that all those prolonged scholarly skirmishes about the autonomy of IndE may have failed to impress the actual speakers:

> The term 'Indian English' is not one that all Indians are comfortable with. Over the years it has borne the connotation of 'bad English'. [...] In any case, the fact remains that, in India, those who consider their English to be good are outraged at being told that their English is Indian. Indians want to speak and use English like the British or, more lately, like the Americans. (Sailaja 2009: 13f.)

If this is an accurate reflection of the IndE speech communities' attitudes and preferences, then Quirk has, after all, carried the day in the Quirk-Kachru debate (cf. Chapter 2.1.2). The "stereotype of a standard language" (Le Page 1988: 30) with respect to English might exercise a stronger hold on the public imagination in India than elsewhere:

> We all know that there is English, spoken by the English-speaking community; that there are various regional and social norms; and that there is at the top of the pyramid an educated written Standard, inscribed on a banner borne through snow and ice by Sir Randolph Quirk. (Le Page 1988: 30)

Mukherjee (2007) has applied Schneider's Dynamic Model to IndE and has analyzed in some detail the conflicting tendencies of exonormative orientation and endonormative stabilization which come to the fore in the quotes above.[14] He has characterized IndE as "marked by an equilibrium, that is, a steady state between progressive and conservative forces" (2007: 173). According to him, "the most important progressive force is the innovation of new forms and structures by Indian users of English" (2007: 174), which contributes to the process of structural nativization of the language. He also sees evidence for "progressive forces at work both at the functional and the attitudinal levels" (2007: 176), that is, a changing attitude towards increasing acceptance of endonormative models. However, it is precisely these two latter levels where conservative forces are most prominent. In line with many other scholars, Mukherjee points to "many Indian speakers' self-critical attitude toward the local form of English that they themselves use" as "perhaps the strongest conservative force at work" (2007: 177). Given these conflicting tendencies, Mukherjee proposes the label "semiautonomous variety" for IndE:

> A semiautonomous variety of English is a nonnative variety that takes over to a very large extent – and includes – the common core of established native varieties of English, but it is also characterized both by interference (that is, forms and structures which can be traced back to speakers' L1s) and by L2-internal creative autonomy (that is, by a potential for the development of new forms and structures in English as a second language). A semiautonomous variety like present-day Indian English can thus be characterized as endonormatively stabilized but showing some aspects of ongoing nativization. In a wider setting, I would hypothesize that any phase 4 variety, that is, a New English variety that is largely endonormatively stabilized without being the dominant native language in the region at hand, can be viewed as semiautonomous in this particular sense – an admittedly speculative claim that will warrant further investigation. (2007: 182).

By way of conclusion, Mukherjee turns back to the Dynamic Model:

> it seems that Schneider's model is more appropriate for settler-strand-dominated varieties such as American and Australian English than for indigenous-strand-dominated varieties such as Indian English. (2007: 173f.)

In essence, Schneider's model is explicitly directed at capturing the social conditions for an autonomous speech community to develop. Once this process is more or less completed, all further steps towards standardization are taken for granted. The notion or process of endonormative stabilization seems to rely on a concept of standardization as envisaged e.g. by Haugen (1966):

14. Mukherjee's article appeared in between Schneider (2003), which outlined the theory of the Dynamic Model but made no reference to IndE, and Schneider (2007), where IndE was included among the case studies.

> In a wider setting, any standard form of any variety of English can be viewed – in a very uncontroversial way – as the result of a standardisation process, which according to Haugen (1996) [sic] includes four aspects: (1) selection of a variety; (2) codification of the variety in terms of dictionaries, grammars, etc.; (3) the elaboration of the selected variety in terms of a wide range of functions; (4) general acceptance of the selected variety as the standard form. This process did not only take place in Britain, where it resulted in Standard English as discussed before, but in many other native-speaker communities, e.g. in American English and Australian English. Also, new standards have emerged – or are about to emerge – in many former colonial territories in which English is used as an "institutionalised second-language variety". (Klippel & Mukherjee 2007: 303f.)

Basically, Haugen's model of standardization describes how a variety of a specific language becomes the standard, in other words, how one of the available norms of a speech community becomes the *only* norm. In order to conceptualize this process and to evaluate the prospects for the emergence of a Standard Indian English, we need to look closer at "norms" and "standards".

2.4.1 Norms and the (post)colonial speech community

We have seen in Chapter 2.2 that the notions of "norm" and "speech community" are intimately related, perhaps even to the point of circularity: a speech community may be defined by its adherence to a set of shared norms, and these shared norms in turn constitute the speech community. Both Koch (1988) and Le Page (1988) have pointed out the necessity to distinguish at least two aspects when referring to "norm". Koch's "descriptive norm" ($norm_d$) (1988: 330) and Le Page's "standard in the sense of norm" (1988: 32) are part of the overall language, but hold only for specific varieties within that language. Prior to standardization, there is no hierarchy attached to the different available descriptive norms; each variety within the language has its own $norms_d$ and its range of functions. Standardization essentially means that one of the available $norms_d$ extends its range, such that, eventually, this descriptive norm becomes coextensive with the prescriptive norm ($norm_p$) for the language as a whole. Once the standard is in place, the standard variety serves as the reference variety, and other varieties are demoted to the status of dialects.

Standardization according to Haugen's model is then typically a process which arises out of the speech community, at least in the initial stages concerned with the selection of a norm. Le Page's characterization of descriptive norms probably applies to the collective process of selecting a norm as well: "A standard in the sense of a norm comes about through focusing. It is likely to be a largely unconscious process" (1988: 32). Further steps beyond the selection of a norm such as codification and elaboration may be subject to official planning, for example by designated

language academies. The history of the standardization of English shows, however, that codification may proceed without a central authority or any other form of institutionalized language planning endeavours (cf. Nevalainen & Tieken-Boon van Ostade 2006).

Mesthrie (2006a) has raised an important point in this respect: when speakers of English began to set up colonies around the world, the process of standardization of the language had only just begun. Further, many of the early settlers, including missionaries and others who taught English to the natives, spoke varieties of English that were both geographically and socially far removed from what was to emerge as the standard:

> For historical veracity we need to keep in mind the following: (a) that Standard English of the period of exploration, trade and colonisation was slightly different from English in the twentieth and twenty-first centuries; and (b) that Standard English was not the only input in the formation of New Englishes. The superstrate was also shaped by sailors, soldiers, adventurers, hunters, missionaries, tradespeople, indentured workers, plantation owners, overseers, settlers, schoolteachers, and – in some island and coastal contexts – divers and whalers. This was a rather varied input, that reminds us that the notion of a TL [target language] is an idealisation. More often – and certainly outside the classroom – the TL was a varied and 'moving' target. (Mesthrie & Bhatt 2008: 188f.)

Mesthrie & Bhatt then argue "for greater historical continuity between the early modern English superstrate and its new English colonial offspring than is generally appreciated" (2008: 198f.). However, I would claim that this point is largely irrelevant for the development of IndE for two reasons. One reason has to do with the specific history of English in India, another is of a more principled nature. Firstly, we have seen in Chapter 2.1.4 that the spread of English in India to any significant extent has to be dated to the 19th century (cf. also Chapter 3.3.), well after the Early Modern English period. Secondly, the nature of the target variety of English is much more important in settlement colonies such as the US than in exploitation colonies such as India.[15] Typical of exploitation colonies is the following scenario:

> The European lexifier was introduced in a scholastic form to a local elite of colonial and missionary auxiliaries only, essentially for them to serve as a managerial group sandwiched between the colonizers and the colonized. In the long run, the indigenization of English in such communities resulted from its appropriation by this group, under decreasing influence from native speakers, with the indigenous elite expanding its range of uses to new, internal communicative functions, originally unintended by the Europeans. (Schneider 2007b: 24f.)

15. For the difference between settlement and exploitation colonies and its significance for the development of postcolonial Englishes cf. Mufwene (2001: 205), Schneider (2007b: 24f.).

By contrast,

> [s]ettlement colonization led to interaction between several varieties of a European source language, integrating speakers of different backgrounds, [...] it is characteristic of such developments that variants of the lexifier were adopted as the vernacular by coexistent communities. (Schneider 2007b: 25)

Schneider then claims that these two different scenarios ultimately have no bearing on the common trajectory of postcolonial Englishes:

> in the long run these settlement and transmission types, important as they are, are not prime determinants of the outcome of the process of new dialect emergence. How and why two groups were brought together and what their relationship was like in the early phases of contact turns out to be less important in the long run than the recognition that once the settler group stays for good they will have to get along together, for better or worse. This insight forces all the parties involved in a contact setting to reconsider and rewrite their perceptions of themselves, their social identities – a process with direct linguistic consequences. (2007b: 25f.)

While I fully agree with Schneider that the identity rewritings in postcolonial speech communities are more important in driving the development of a postcolonial English, I would claim that the process of standardization has to be conceptualized differently for Englishes such as IndE, which emerged in exploitation colonies.

English in India has never been a vernacular comparable to English in the US, and it has mostly been acquired in schools and universities rather than being transmitted from one generation to the next.[16] It is probably only now, at the beginning of the new millenium, that something akin to an IndE vernacular is developing. Political and social changes over the last three decades have contributed to the unprecedented rise of a large Indian middle class, urban and cosmopolitan, whose children will frequently grow up with English as part of their overall linguistic repertoire. Thus, the dominance of a taught "scholastic form" of English in India is now complemented by a truly vernacular IndE, as English has become a language of daily interaction for a still small, but growing number of the population. An alternative to Haugen's model of standardization which is better suited to such a scenario will be the topic of the next section.

2.4.2 Spoken and written standards

Haugen's model was developed with the historical development of European vernaculars to national as well as *written* languages in mind, with the processes of

16. The community of Anglo-Indians probably forms the only exception of any significance (cf. Coelho 1997).

standardization and textualization inextricably linked (cf. Haugen 1966: 929). By contrast, the exonormative standard of English in India, though certainly not immutable over the centuries, was not negotiable to the same extent as norms and standards are negotiable in a settlement colony such as the US:

> American ideas concerning the autonomy of their form of English [...] did not influence the emerging Englishes elsewhere. In many colonies the number of locals with a mastery of English was generally much smaller than those still learning it. (Mesthrie & Bhatt 2008: 23)

Standardization always involves reduction of variability, regardless of whether standardization proceeds as a largely unconscious process of norm-negotiation or as deliberate planning. Nevalainen & Tieken-Boon van Ostade provide examples where "change from below eventually won the day" in the choice of specific morphological features in the English system of verb inflection (2006: 294). That is, norms arising out of the spoken language became part of the spoken as well as the written standard. Such a process will not occur when the norms for the written language and partly also for the spoken language have already been set, as is the case with former colonial Englishes and current institutionalized second-language varieties of English.

The model of standardization proposed by van Marle (1997) is particularly well suited to capture the interaction between an exonormative written standard and emerging local spoken standards.[17] Van Marle first contends that

> there is a fundamental difference between standard languages that are exclusively written, and standard languages which, alongside their written form, can be said to have a spoken variety as well. (1997: 13)

According to van Marle, such a division between the spoken and the written language, with only the latter being standardized, was typical of much of Europe up to the 19th century:

> The standard language functioned primarily as a written code, which meant that for centuries it was only a relatively small part of the population which was actively engaged in using it. [...] Typical of this situation is that there is hardly any interaction between the written standard and its related dialects, meaning that the two levels are largely independent of each other. That is, from the moment that the formation of the written standard was a fact, dialect features hardly penetrated into this largely written variety. The influence of the written standard on to the dialects was highly limited as well. (1997: 14)

17. The model has been developed as a description of the linguistic situation in the Netherlands, but can be applied to other contexts as well.

When a spoken standard developed in the 19th century, it did not emerge "from below", growing out of the available competing norms of the speech community, but took its clue from the written standard: "the emerging spoken standard was rooted in the written standard or, more precisely, the written standard gradually came to be accepted as a norm for speaking" (1997: 15). Van Marle continues:

> The fact that the written standard became a determining factor for speaking resulted in a language variety – the spoken standard – which is not as formal as the written standard, but which has as its defining characteristic the fact that *it is directed towards the written language.* That is, it is the written standard which serves as a point of reference for the spoken standard; the more formal the situation is, the more evident the influence of the written language on cultivated speech. This implies that *it is the written language which influences its spoken counterpart*. (1997: 16, emphasis mine)

In India, written Standard British English was disseminated through schools, universities, the administration and the relatively new print media. Krishnaswamy & Burde illustrate how the written standard influenced the spoken mode throughout the history of English in India. They note for the early stages of English in India that "the ostentatious formality in written correspondence, especially in official correspondence, is found even today in the Indians' use of English" (1998: 82). For a later stage, they observe the "extension of the bureaucratic use of English outside the bureaucratic domain" (1997: 107). Sedlatschek's corpus-based study has confirmed that "present-day IndE carries numerous traces of its origin in the brand of educated English evolving in India in the second half of the nineteenth century" (2009: 314).

Once a spoken standard has developed, then the interrelations between the different levels can be visualized as in Figure 2.3.

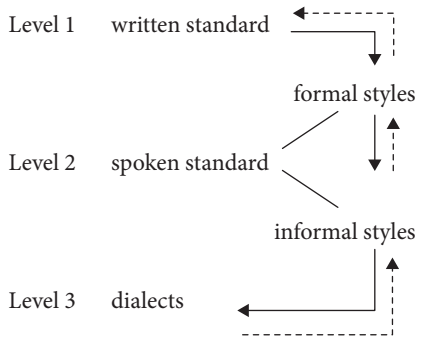

Figure 2.3. "Relationship between a written standard, its spoken counterpart, and the related dialects" (van Marle 1997: 19), non-dotted arrows "indicate basic patterns of influence", dotted arrows indicate "secondary" patterns of influence (*ibid.*)

Van Marle further notes "systematic differences between dialects and their related standard language" (1997: 21). There might be

> Characteristics present in the standard language and absent in the dialects which should be attributed to the process of naïve language engineering that standard languages are subject to. (*ibid*.)

An example would be the prescriptive *it is I* rather than *it is me* (cf. Chapter 4.5), the preferred form in informal spoken and written English. *It is I* never gained ground despite centuries of prescriptivist efforts. Secondly, van Marle lists

> Characteristics absent in the standard language and present in the dialects, whose emergence should be attributed to the greater ease with which natural changes take place (and particularly 'spread') in the dialects. (*ibid*.)

Finally, van Marle refers to the differences arising from the parameters of orality and literacy (cf. Chapter 3.3):

> Characteristics absent in the standard language and present in the dialects, whose emergence should be attributed to the fact that dialects are spoken languages *par excellence*, in which strategies relating to the actual process of speech production are much more influential than in the standard language. (*ibid*.)

Van Marle's model offers a partial answer to the question why IndE so far has not progressed any further towards endonormative stabilization. The exonormative British English standard is still in place for the written mode and still very influential for the spoken mode. The indigenous IndE norm manifests itself at van Marle's third level. The tension between the first level of the exonormative British English standard and the third level of endonormative spoken IndE will then become apparent at the second level, the spoken standard. Chapter 4 will be devoted to an in-depth analysis of the function and distribution of some features of spoken IndE, Chapter 5 will evaluate the findings from Chapter 4 with respect to a model of standardization that acknowledges the interdependencies between written and spoken, exonormative and endonormative standards and norms.

CHAPTER 3

Multilingualism in India

3.1 Introduction

All of the concepts introduced in the previous chapter – New Englishes, Outer Circle, institutionalized non-native variety of English, Postcolonial Englishes, to name but a few – are predicated upon multilingual speech communities in which English plays a prominent role. While it may seem trivial to state this, the discussion in Chapter 2.3 has shown that conceptualizing the relation between multilingualism, language contact and language change is far from trivial. More specifically, if we want to assess the future development of English in India, we obviously need to take not only strictly linguistic aspects, but also socio-historical and other cultural aspects into account. It is, after all, speakers' attitudes towards their own variety that are instrumental for changing its status, not some inherent properties of the variety in question. Indeed, Schneider's Dynamic Model presupposes that "English is being accepted and appropriated by IDG strand populations, and that in its indigenized form it becomes an identity carrier" (Schneider 2007b: 64), that is, the process of appropriation is more important for the emergence of a New English as a distinct variety than the degree of linguistic 'distance' to other varieties caused by nativization.

The establishment of American English may serve as an example for successful appropriation: English as used in America by Americans gradually became emblematic as a marker of a new national identity, setting the new independent state off from its former colonial ruler.[1] The new linguistic self-confidence is acted out by emphasizing difference, more or less irrespective of actual linguistic differences: the fact that the morphosyntactic differences between AmE and BrE are negligible and that mutual intelligibility is never compromised is of secondary concern here. Language and language use is not a decontextualized issue for speakers, neither on a national nor on a more local level, and issues of language use and language choice are particularly complex in multilingual societies with a colonial heritage.

1. For details see Schneider (2007b: 282–294). Schaefer (2010) discusses the American case in the larger context of language ideologies and linguistic nationalism.

3.2 Typological survey

Chapter 3.2 will be mainly concerned with the linguistic typology of South Asia in several distinct but intertwining respects. I will first describe briefly the patterns of multilingualism in modern India, mainly with reference to the distribution of language families, individual languages and their speakers. I will then highlight some convergence features within the South Asian *sprachbund*, focussing on those aspects which will be taken up in later chapters. The notion of *sprachbund* is typically applied to shared phonological and morphosyntactic features between genetically unrelated languages; for South Asia, the concepts "discourse area" and "sociolinguistic area" have been suggested. Their significance for this study will be discussed in Section 3.2.3 and 3.2.4.

Chapter 3.3 briefly examines the position of English in the Indian communicative space. Although intuitively obvious, the notion of "communicative space" entails a specific theoretical perspective which will be spelled out here. The aim of this section is to trace and assess the gradual establishment of English as an Indian language. As a whole, the chapter will provide the necessary background for an evaluation of the IndE language contact scenario, both in terms of specific linguistic features and with respect to an emerging IndE standard.

3.2.1 Patterns of multilingualism

To determine the exact number of languages that are spoken within India alone, let alone South Asia, is fraught with the usual difficulties. The first person to undertake a taxonomy of the languages spoken in (British) India was George Abraham Grierson, and his massive *Linguistic Survey of India,* which was begun in 1894 and took decades to complete (the first edition was published from 1904 till 1928), continues to serve as the basis for all further efforts in the field.[2]

The compilers of the Indian census, which has been carried out every 10 years since 1881, have found wildly varying answers to the question of how many languages are spoken in India.[3] The *Linguistic Survey of India* originally named 179 languages spoken in British India. The 1961 census, the first census following India's independence, famously listed 1652 "mother tongues", whereas the 2001 census recognizes 122 languages, with 234 mother tongues subsumed under these

2. A new language census is in preparation at CIIL, Mysore.

3. The first systematic and comprehensive Indian census was actually carried out from 1865 to 1872; the 1881 census was the first to be undertaken simultaneously across the country. It was also the first to include the category "mother tongue".

languages.[4] Jain's account of how the census compilers bridge the gap between speakers' own perception of 'mother tongue' and their classification schemes is instructive (2003: 47):

> There is no agreement on the number of languages and dialects spoken in the subcontinent. The Census of India recorded 845 languages spoken in 1951, 1652 in 1961, 105 in 1971, 105 in 1981 and 114 in 1991. It is interesting to see how the census arrived at this number in 1991. The respondent was given the freedom to name his mother tongue which the census enumerator recorded faithfully. The number of raw returns of mother tongue came to 10,400. This number was subjected to linguistic scrutiny which resulted in 1576 rationalized mother tongues and 1796 names which were treated as 'unclassified' and relegated to 'other' mother tongue category. Linguistic methodology was further applied to the 1576 mother tongues to arrive at the number 114 (Census of India 1991, Language: 8). This number, however, does not include the languages spoken by less than ten thousand speakers, a practice which the census has followed since 1971.

The process of the reorganization of states along linguistic boundaries, which began shortly after Independence despite resistance from Nehru and his first Union government, has continued to the present day (cf. Lange 2010). The eighth Schedule of the Indian constitution originally gave official recognition to 14 languages. By 2003 the list included 22 languages. Even if there are no immediate benefits directly associated with being "on the list", speech communities have often lobbied extensively to have their language promoted to "official" status by being included in the eighth Schedule. Agnihotri (2007b: 194) called it

> a stroke of "raw genius" [...] to leave the list of languages under the VIIIth Schedule "Languages" open. This allowed a speech community to have its language included in the list when its members felt the need to assert their identity.

Even though the definition and demarcation of languages spoken in India continues to be a site for socio-political conflict, the typological boundaries between language families that are present in India are well-established (cf. Comrie 1998, Singh 2003, Subbarao 2008). The four major language families of India (Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman) differ notably in terms of the number of speakers and in terms of the number of languages belonging to the family: while the Indo-European language family has the highest number of speakers, the Tibeto-Burman language family covers the largest number of individual languages (Asher 2008: 33).

4. Only languages with more than 10.000 speakers are included in the census. For the ideological underpinnings of the distinction between "language" and "mother tongue" in the census cf. Lange (2010).

Table 3.1. Typological distribution of the 122 scheduled and non-scheduled languages of India in the 2001 census (http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/statement9.htm)

Language families	Number of Languages	Persons who returned the languages as their mother tongue	Percentage to total population
1. Indo-European			
a. Indo-Aryan	21	790,627,060	76.87
b. Iranian	2	22,774	0.00
c. Germanic*	1	226,449	0.02
2. Dravidian	17	214,172,874	20.82
3. Austro-Asiatic	14	11,442,029	1.11
4. Tibeto-Burmese	66	10,305,026	1.00
5. Semito-Hamitic**	1	51,728	0.01
Total	122	1,026,847,940	99.83

* i.e. English

** This language family is represented by Arabic in the 2001 census, it is generally not included in typological surveys of India.

Indo-European and Dravidian languages can be roughly assigned to the north and the south of India, respectively. Tibeto-Burmese languages are clustered along the northern and eastern border areas, whereas Austro-Asiatic languages belonging to the Munda branch are spoken in eastern, north-eastern and central parts of India (Asher 2008: 42f., cf. also Breton 1997, Singh 2004). However, it has to be kept in mind that typological classifications are in reality never as clear-cut as they appear on paper, in particular not at the level of grassroots multilingualism, as Gumperz and Wilson (1971) have famously demonstrated. Boundaries between individual languages of the same family are also not as clear-cut as one might expect, given the Western experience of separating languages along national lines:

> The entire Indo-Aryan realm (except for Sinhalese) constitutes one enormous dialectal continuum [...]. The speech of each village differs slightly from the next, without loss of mutual intelligibility, all the way from Assam to Afghanistan. Cumulatively the differences are very great, but where to draw the dialect, let alone the language, boundaries? (Masica 1993: 25)

In India, traditional patterns of multilingualism combine with more recent patterns of migration, to the effect that there are sizable minorities within larger speech communities in all Indian states, and especially in the metropolitan areas. Table 3.2 depicts the extent of linguistic heterogeneity for five randomly selected Indian states: even though the speakers of the states' dominant regional language are clearly in the majority, each state's linguistic heterogeneity remains remarkable.

Table 3.2. Distribution of 10,000 persons by language in selected Indian states and union territories from the 2001 census, adapted from http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement3.htm. 'n' stands for 'negligible'

	Delhi	Maharashtra	Nagaland	Kerala	Tamil Nadu
Assamese	5	n	93	n	n
Bengali	151	32	326	1	1
Bodo	n	n	27	n	n
Dogri	5	n	3	n	n
Gujarati	33	239	3	6	32
Hindi	8100	1104	315	8	30
Kannada	8	130	2	26	168
Kashmiri	15	1	n	n	n
Konkani	1	68	n	19	1
Maithili	62	4	4	n	n
Malayalam	66	42	22	9676	89
Manipuri	1	n	38	n	n
Marathi	19	6889	8	10	10
Nepali	32	7	189	1	1
Oriya	21	10	22	1	1
Punjabi	715	28	8	1	1
Santali	n	n	1	n	n
Sindhi	31	73	n	n	1
Tamil	67	55	8	188	8943
Telugu	20	145	7	15	565
Urdu	632	713	4	4	151
Sched. L.	9984	9540	1083	9959	9995
Nonsched. L	16	460	8918	41	5

This table taken together with Table 4.8 in Chapter 4.1.3.4 depicting speakers' bi- and trilingualism indicates the degree of individual and societal multilingualism across India, and there is no evidence that this situation is likely to change.

In the following, I will briefly consider one of the possible outcomes of extensive multilingualism and language contact, namely convergence.

3.2.2 South Asia as a *sprachbund*

In all discussions around *sprachbund* phenomena, South Asia appears as a prominent example (alongside the Balkan area).[5] The outcome of language contact

5. For a critical discussion of the notion of *sprachbund* see Matras (2009: 265–268).

especially between the Dravidian and the Indo-Aryan languages has been extensively studied ever since the pioneering works by Emeneau (1956) and Masica (1976).[6] It should be kept in mind, however, that convergence in India is not limited to the Indo-Aryan and Dravidian languages:

> it is now widely acknowledged that all the four original families (Indo-Aryan, Dravidian, Austric, Tibeto-Burman) of languages in contact on the subcontinent today share more specific features among themselves than any one does with an external member of the family [...]. Such a concept of linguistic area means that contact has been even more prevalent than genetic affiliation: the links of blood, so to speak, were superseded by links of neighbourhood. (Montaut 2005: 92)

For Masica, two quite unrelated features mark South Asia as a linguistic area: the "two essentially South Asian features" (2001: 254) are retroflex consonants and the "dative-subject" experiencer constructions, defined as "the generalization of a marking of non-volitionality [...] by means of the consistent special case marking of Experiencers" (2001: 257). While these two features taken together may serve to uniquely delineate the South Asian convergence area, another feature which is highly relevant in the present context extends far beyond South Asia (cf. Masica 1976: 30): Masica lists SOV word order as one "of the most striking areal characteristics of South Asia" (2001: 240). The only exceptions on the Indian subcontinent are Kashmiri (Indo-Aryan) and Khasi (Austro-Asiatic), which both have basic SVO word order (cf. Ebert 2001: 1530).[7] Indian languages are further generally postpositional; the auxiliary follows the verb, and modifiers such as adjectives, possessives, or numerals precede the noun. Masica is careful to point out that these latter features frequently co-occur with basic SOV order, but not necessarily so:

> *They are therefore best treated as independent variables*. They cluster together or fail to do so in an areally significant manner. (2001: 241, emphasis in the original)

Another noteworthy areal syntactic property is the general absence of definite articles (Masica 2001: 245), which has been called upon to explain the over- or under-use of articles in IndE (cf. Sharma 2005). The list of South Asian areal features is of course much longer (cf. Masica 1976, 2001), however, what is important in our context is, first of all, the sheer extent of areal convergence, which testifies to prolonged contact under conditions of stable multilingualism, which in turn is

6. The state of the art in South Asian areal typology was the topic of a special volume of the *Yearbook of South Asian languages and Linguistics 2001*: "South Asian Languages: Contact, Convergence and Typology".

7. Masica tentatively attributes the exceptional status of both languages to "retention of genetic heritage [...]. It can be noted that the Khasi Hills and the Kashmir Valley share a notable degree of geographic and historic isolation" (1976: 39).

likely to defy attempts at formulating clear-cut constraints on contact-induced language change. Masica, who is very sympathetic to Thomason and Kaufman's borrowing scale (cf. Chapter 2.3 above), nevertheless notes:

> All of this has been concerned with contact relations between two *languages only* – i.e., with the simplest cases. When it comes to *linguistic areas*, Thomason and Kaufman can only say that they are notoriously "messy". Prolonged and complex contacts are involved among more than two languages, and more than one of the above-described processes is likely to have been at work. That is, there has probably been both shift and borrowing, and sometimes even creolization. Moreover, in a zone of prolonged contacts (such as South Asia), there may be successive layers of influence, and earlier influences may partly be obscured by later ones. Finally, not merely bilingualism, but widespread *multilingualism* seems to be typical of linguistic areas (and in fact of much of human sociolinguistic history). In such situations there is no one source of diffusion of features, but many centers, now here, now there. (Masica 2001: 234, emphasis in the original)

The other important point concerns further ramifications of the almost universal SOV word order, to which I will now turn.

3.2.3 South Asia as a discourse area

In their paper "Discourse level evidence for South Asia as a linguistic area" (1991), Moag & Poletto proposed four features evident at the level of discourse organization which they consider to be distinctive for the South Asian *sprachbund*. Two of their features explicitly deal with written language: the first feature concerns the textual organization of paragraphs in expository prose, which are said to have a "spiral structure" (Moag & Poletto 1991: 232). They further refer to a "universality of communicative goals" in South Asian art forms including writing, which they characterize as "(1) the supercedence of mood over information; and (2) the primacy of form over content" (1991: 234). Two problems arise with these two proposals. Firstly, both features are stated in such broad – if not to say vague – terms that they are hardly amenable to empirical testing. Secondly, it is doubtful whether linguistic features that are explicitly related to the written medium can at all be used to elucidate the notion of *spachbund;* written language, and particularly poetic written language, belongs to a highly specific register far removed from everyday language use, especially in societies in which literacy cannot be taken for granted.

With their third feature, Moag & Poletto return to the spoken realm: they note the widespread use of zero anaphora for establishing text cohesion and for "marking the distinction between old versus new information" (1992: 238):

> With respect to anaphora, the two language families [Indo-Aryan and Dravidian] seem to share a convention in connected discourse for zeroing out most, and frequently all, old information which is capable of pronominalization. [...] zeroing of the subject pronoun is the unmarked condition in dyadic exchanges in South Asian languages. (1991: 239)

Their final feature concerns the pragmatic functions of the well-known phenomenon of compound or serial verbs in South Asian languages (cf. Emeneau 1956: 9), namely "deixis" (Moag & Poletto 1991: 241) and "conveying speaker attitude" (1991: 243). In our context, feature three, the widespread use of zero anaphora, is the most important. With respect to Moag & Poletto's observations, it remains an unsolved issue whether the notion "linguistic area" can be made more coherent by including shared properties of discourse organization.

Ever since Li & Thompson published their classic paper on the typology of subject-prominent versus topic-prominent language (1976), zero anaphora belong to the prime criteria for establishing that a language is topic-prominent. Topic prominence is strongly associated with verb-final word order, and the preceding discussion has shown that basic SOV order extends beyond South Asia and is thus not a *sprachbund* feature in the strict sense, at least not in isolation. Moag & Poletto might not have been aware of the distinction between subject-prominent and topic-prominent languages suggested by Li & Thompson, as they make no reference to the text and do not discuss zero anaphora together with basic word order. It does not come as a surprise, then, that their suggestion for establishing a South Asian "discourse area" has not been followed up in the literature. The notion of topic-prominence in relation to zero-anaphora, however, will prove relevant in subsequent chapters, and I will now turn to a brief discussion of associated concepts and terminology.

Li & Thompson (1976) have proposed a typological classification of languages according to the relative prominence of subject or topic in encoding grammatical relations. In subject-prominent languages such as English, the syntactic notion of subject is detached from the discourse-pragmatic notion of "topic" and takes precedence: the sentence structure is subject-predicate. This does not prevent subjects from being co-extensive with topics, nor does it mean that subject-prominent languages are completely devoid of topic-comment structures. It just means that the syntactic category of subject is structurally more important than the categories of information structure such as topic and comment. The distinction may best be illustrated with English sentences with a "dummy" subject (Li & Thompson 1976: 467):

(3.1) a. It is raining.
b. There is a cat in the garden.

Subject-prominent languages with a strict word order such as English require every sentence to have a subject, even if it is nonreferential or semantically empty. Topic-prominent languages, on the other hand, mark the topic but not necessarily the subject:

(3.2) Sakana wa tai ga oisii
fish TOP. red snapper SUBJ. delicious
"Fish (topic), red snapper is delicious." (Japanese)

(3.3) Nèike shù yèzi dà
that tree leaves big
"That tree (topic), the leaves are big." (Mandarin)
(Li & Thompson 1976: 468)

In the Japanese example (3.2), both the topic and the subject are marked by particles; in Mandarin, the topic is marked by its sentence-initial position. Li & Thompson refer to examples such as these as "double subject" constructions, the prototypical instances of topic-prominent syntax:

> Such sentences are, of course, the clearest case of topic-comment structures. First, the topic and the subject both occur and can thus be distinguished easily. Second, the topic has no selectional relationship with the verb. Third, no argument can be given that these sentences could be derived by any kind of "movement" rule from some other sentence type. (1976: 468)

The structure of such "double subject" constructions would then be:

(3.4) $\underline{NP_1}$ $\underline{NP_2}$ V
topic comment (Li & Thompson 1976: 485)

Topic-prominence is strongly correlated with verb-final word order, as (3.4) already suggests. Other diagnostic criteria for topic-prominent languages put forward by Li & Thompson include the absence of dummy subjects as already mentioned, the lack or marginal use of passive constructions, and zero anaphora (1976: 467).[8] While it is far beyond the scope of this study to assess the relative degree of topic-prominence of individual Indian languages, even a cursory look at grammars and relevant research confirms that Indian languages are much less subject-prominent than English. For Hindi, Kachru notes (1987: 72):

> a closer correspondence between semantic and syntactic grammatical roles that nominal constituents have in a sentence [...]. Many of these characteristics of Hindi-Urdu are shared by not only the other Indo-Aryan but also the Dravidian and other languages of India.

8. Cf. Krishnamurti (2003: 466–68) for Dravidian languages and Junghare (1988) for Indo-Aryan languages.

Word order in Indo-Aryan as well as Dravidian languages is relatively free and allows more or less any constituent to occupy the sentence-initial topic position.[9] The Marathi examples illustrate a locative and a topic moved out of a subordinate clause to the sentence-initial topic position (Pandharipande 1997: 252):

(3.5) gharātʃyā tʃhaprāwar to basiā hotā
house-POSS roof-on he sit-PST-3SM was
'On the roof of the house, he was sitting.'

(3.6) tyātʃī bahīṇ he bara dzhāla
he-POSS-3SF sister this good happen-PST-3SN
kī amerikelā gelī nāhī
comp America-ACC go-PST-3SF NEG
'His sister, it was good that she did not go to America.'

Similarly, the examples from Malayalam feature a direct object and another locative promoted to topic position (Asher & Kumari 1997: 184):

(3.7) ravi varmayuṭe citram oru ameerikkakkaraan vaaŋŋi
Ravi Varma-GEN painting an American buy-PAST
'Ravi Varma's painting, an American bought it.'

(3.8) kooṯṯayatteekə ŋaan kaẓiŋŋa aaẓca pooyirunnu
Kottayam-ALL I last week go-PERF$_1$-PAST
'Kottayam, I went there last week.'

Additionally, Indian languages have optional topic markers, as in the following example from Hindi:

(3.9) **aj** (**to**) həm ʈenis zərūr k^{h}elēge
today (PTCL) we tennis certainly play.FUT.M.PL
'*Today* we will definitely play tennis.' (Kachru 2006: 246)

According to Kachru (2006: 246), the clitic particle *to* marks the adverbial *aj* 'today' as the theme in the answer to a question like "What will you do today?" Both of these pan-Indian structural properties, i.e. the freedom of word order which allows practically any constituent to become the sentence topic, and the availability of topic-marking clitics, will have some bearing on the discussion of IndE discourse-pragmatic word order in the following chapter.

The notion of topic-prominence has also been extended from its application to a typological parameter of syntactic configurationality to a specific mode of

9. Cf. Asher & Kumar (1997: 183) for Malayalam, Pandharipande (1997: 246–253) for Marathi, Kachru (2006: 245–251) for Hindi.

discourse organization: topic-prominence has been associated with a more "pragmatic" rather than a "syntactic" discourse structure (cf. Givón 1979a, b) and further with a possible grammaticalization trajectory:

> subject and topic are not unrelated notions. Subjects are essentially grammaticalized topics. (Li & Thompson 1976: 484)

As we shall see, a more pragmatic discourse mode has also been ascribed to New Englishes generally due to their origin in language-learning situations (cf. Mesthrie 1992: 123f. and Chapter 4.3.6 below), which leaves us with competing motivations for many of the phenomena to be discussed in the next chapter.

3.2.4 South Asia as a sociolinguistic area

While the concept of *sprachbund* or "linguistic area" generally refers to shared structural features of languages, the concept "sociolinguistic area" draws attention to "features of language *use* that cluster in areal relationships" (Ferguson 1996: 85, emphasis mine). Jean d'Souza has extended the areal-typological perspective on South Asia by integrating sociolinguistic aspects, and she was the first to provide a definition of "sociolinguistic area" as

> one in which languages belonging to several different families converge as a result of their daily interaction and because they are called upon to express certain shared social and cultural experiences and adhere to sociocultural norms accepted by the members of the society at large. Such an area is marked by widespread bilingualism or multilingualism and by the fact that common social, cultural and linguistic factors have influenced the languages of the area. A sociolinguistic area, therefore, is characterized by diverse social groups and diverse language families both of which in the course of time begin to share a grammar [of] culture and a grammar of language. (d'Souza 1988: 159)

Seen from one perspective, the two concepts "linguistic area" and "sociolinguistic area" are independent, even dichotomous: the main orientation of areal typology is diachronic, with the aim of classifying languages through the historical reconstruction of the structural outcome of language contact. The concept of *socio*linguistic area, on the other hand, implies a more synchronic perspective in identifying shared patterns of current language use; that is, the dichotomy "diachronic – synchronic" is accompanied by the dichotomy "language structure – language use". However, the undeniable fact that "convergence areas are no doubt most closely related to cultural areas" (Masica 2001: 217) provides the starting point for a more integrated perspective on the two approaches: "the sociolinguistic area concept subsumes the linguistic area concept because a linguistic area is the result of the prior formation of a sociolinguistic area" (d'Souza 1991a: 297).

Ferguson (1996) has taken up the concept and discussed some features that he sees as constitutive of a South Asian sociolinguistic area. His features range from macro-linguistic issues pertaining to language use and language choice across South Asia to more "specific communicative subsystems" (1996: 86). I take Ferguson's features as a starting point for further discussion, as they collectively represent a broad perspective on the Indian communicative space that other scholars have treated separately. Khubchandani (1991), for instance, uses the term "sociolinguistic area" with respect to India to come to terms with the plurality and functional heterogeneity of languages and language identities across the subcontinent. D'Souza herself is more interested in the actual linguistic reflexes that emerge from the sociolinguistic *sprachbund*, and I will come back to her research below.
Ferguson's first three features delineate aspects of the South Asian communicative space that are integral to its sociocultural heritage and thus relevant both synchronically and diachronically. The first feature concerns the topic of Section 3.2.1 above, namely multilingualism: "No other region of the world has had such a long-continued pattern of socially accepted, governmentally institutionalised multilingualism" (Ferguson 1996: 86). Two additional features are linked to literacy: the ancient tradition of literacy and the equally ancient "official recognition of and public use of different systems of writing" (1996: 87). Two further macrolinguistic features refer to patterns of sociolinguistic variation that are unique to South Asia: Ferguson argues that "caste membership may be the single most important parameter of social dialect variation in South Asia" (1996: 88), overriding all other parameters of variation familiar from sociolinguistic research. Another feature which has to do with variation concerns register variation, more specifically the multi-layered "South Asian diglossias" (1996: 89).

With three final features, Ferguson turns to shared linguistic features at the interface of syntax and pragmatics: kinship terms, forms of address and politeness formulas (1996: 90–93). It is in this area that the notion of a South Asian sociolinguistic area becomes most important for the purposes of the present study. D'Souza (1988) treats pronouns of address and kinship terms as a subgroup of pan-South Asian devices for expressing politeness, more specifically as belonging to a set of morphological/lexical means to indicate politeness. What distinguishes d'Souza's approach from Ferguson's with its emphasis on areal typology, albeit an extended concept of areal typology, is that she explicitly relates the South Asian "grammar of culture" to the emergence of an indigenized variety of English: "English in South Asia is both adapting to the local grammar of culture and influencing it" (1988: 168). D'Souza traces the notion "grammar of culture" back to an article by Bright (1968), noting that the concept "was not widely accepted at the time" (d'Souza 1988: 160):

> 'Grammar of culture' is used here to mean the acceptable possibilities of behaviour within a particular culture. This includes notions of the kind of behaviour that is *appropriate* or *expected* in a given context. Since the use of language is included within the 'acceptable possibilities of behaviour', some correlation may be found between socio-cultural factors and their linguistic manifestations.

The field of linguistic politeness provides examples for the influence of the "grammar of culture" working both ways: it is well known that the frequent use of *please* and *thank you* as in BrE and AmE is misplaced in the South Asian context, where overt expressions of thanking are considered unduly formal and would not be used within the family or with friends, because they create social distance. However, speakers of IndE are likely to generalize the exonormative pragmatic rules to the Indian context and "use 'thank you' more frequently, in more situations and to more people, *even where the grammar of culture would deem its use incorrect*" (d'Souza 1988: 168, emphasis in the original). Sailaja even notes: "In modern times, *thank you* and *thanks* are used by bilinguals even if they are conversing in an Indian language. The same is true of *please* and *sorry*" (2009: 94, fn. 5). Conversely, South Asian interactional strategies are grafted upon Indian English, as d'Souza illustrates: an expression meaning 'a little' is common in most South Asian languages and is used "to make requests or to help soften imperatives" (d'Souza 1988: 167), as in the following example from Hindi:

(3.10) zarā yahā ānā
a little here come
'please come here' (1988: 167)

Interestingly, then,

> Indian English makes use of the word 'just' in the making of requests, or as a means of minimizing an action, as a substitute for *zarā*. Thus, one often hears people say 'just I will go and come' when requesting permission to leave, or 'just one doubt' when they want to ask a question. (d'Souza 1988: 168)

Sailaja, writing twenty years later, confirms that this "use of *just* is extremely common and is completely acceptable in standard IE as well" (2009: 88). IndE *just* in this sense, then, constitutes a prime example of a pragmatically motivated calque where the source language is not one particular language, but the shared pan-South Asian "grammar of culture". The individual linguistic phenomena discussed in Chapter 4 will afford ample opportunity to return to the concepts spelled out above.

3.3 English as an Indian language

3.3.1 Introduction

When Schneider noted that "English in India [...] is a topic marked by never-ending paradoxes" (2007b: 161), he might well have had the representation of English in the Indian Constitution in mind. Article 343 proclaimed Hindi the official language of the Indian Union, with English as the associate official language for another 15 years.[10] Further articles focussed on the restriction of the use of English for official purposes in the future (§ 344), and, concomitantly, on supporting the elaboration and spread of Hindi (§ 351) (*Constitution of India* 149–152). However, such declarations of intent were not devoid of a certain irony:

> The constitution adopted in 1949 was in English. The Hindi version was authorized, after a long delay, in 1987. In case of any divergence of meaning between the two texts, the Hindi version would be subject to revision. (Dasgupta 2003: 31)

The irony – or paradox – in this case can be seen from two different perspectives. On the one hand, the British colonizers who imposed their language together with their legal and administrative system upon the colonized also unwittingly gave the subjected people the means to overthrow colonial rule. The logic of this process has been most convincingly captured by Anderson, whose Marxist description of the "output" of the colonial education system probably applies to most key figures in the Indian independence movement:

> The expansion of the colonial state which, so to speak, invited 'natives' into schools and offices, and of colonial capitalism which, as it were, excluded them from boardrooms, meant that to an unprecedented extent the key early spokesmen for colonial nationalism were lonely, bilingual intelligentsias unattached to sturdy local bourgeoisies. (2006: 140)

Thus, the fact that the Indian constitution was written in English can be seen as the culmination of the long process of turning colonial knowledge against the colonizers.[11]

On the other hand, having to use the colonial language for a document which announces India's sovereignty might be seen as a partial defeat, as a sign that independence had not been fully accomplished. Gandhi himself, who did not live to

10. After riots in South India, any temporary restrictions on the status of English as associate official language were dropped (cf. Lange 2010).

11. For overviews of the ideology, production, and application of colonial knowledge in India, specifically with respect to language cf. Cohn (1986: 16–55), Washbrook (1991), Mitchell (2005), Trautmann (2006).

witness the constitution coming into effect, had always linked language issues to the struggle for liberation. His assessment of the role of English is worth a longer quote:

> English is a language of international commerce, it is the language of diplomacy, it contains many a rich literary treasure, and it gives us an introduction to Western thought and culture. For a few of us, therefore, a knowledge of English is necessary. They can carry on the departments of national commerce and international diplomacy, and for giving to the nation the best of Western literature, thought, and science. That would be the legitimate use of English. Whereas today English has usurped the dearest place in our hearts and dethroned our mother tongues. It is an unnatural place due to our unequal relations with Englishmen. The highest development of the Indian mind must be possible without a knowledge of English. It is doing violence to the manhood and specially the womanhood of India to encourage our boys and girls to think that an entry into the best society is impossible without a knowledge of English. It is too humiliating a thought to be bearable. *To get rid of the infatuation for English is one of the essentials of swaraj*. (Gandhi, "Young India", 9.2.1921; emphasis mine)[12]

The issue whether "Decolonising the Mind" (cf. wa Thiong'o 1994) necessarily involves getting rid of English is still hotly disputed and no longer tied to any current process of gaining political independence. However, I would like to discuss the significance of the fact that the Indian constitution *had to* be written in English from a much narrower theoretical perspective. English was established in the Indian communicative space in specific, highly compartmentalized functional realms, most of them originally associated with the written medium. In the following, I will elaborate on the notion of "communicative space" and its usefulness for coming to terms with the Indian multilingual scenario.

3.3.2 The Indian communicative space

Many histories of English in India are available (e.g. Kachru (1994), Krishnaswamy & Burde (1998: 79–137), Schneider (2007b: 161–173), Sailaja (2009: 95–112), Sedlatschek (2009: 8–34)) which generally seek to establish different phases in the spread of English on the Indian subcontinent. Most authors begin their narrative of the rise of English in 1600, the date the East India Company was granted the Royal Charter for trade with India. Such timelines seem to suggest that English in India has a long, unbroken and linear tradition of growth and spread, rising more or less inevitably from one trader's language among many to the administrative language of the colonial powers and their subjects, and finally to the associate official language of independent India. As many authors recognize, the actual step from the haphazard

12. *Swaraj*: 'self-rule', i.e. independence.

use of English in India to the firm establishment of English within the Indian communicative space was taken much later. In Krishnaswamy & Burde's terms:

> The link between English education and the British bureaucracy was established during this period [1813–1857]. Print-media and creative writing in English started earlier than formal English education in India. Moreover, most of these developments took place at the three presidency cities – Bombay, Calcutta and Madras; the locale was typically urban. [...] English education and bureaucracy emerge almost simultaneously and become a decisive combination for the use of English in India. (1998: 88f.)

The concept of "communicative space" is similar to the notion of "linguistic ecology", which has been introduced by Mufwene (2001) and has subsequently found its way into many studies in the field of New Englishes (e.g. Lim 2007, Schneider 2007b, Ansaldo 2009). Mufwene's concept incorporates a unified account of language evolution and abolishes the need to distinguish between different types of language change. He specifically challenged the view that the development of Pidgins and Creoles is unlike any other language with 'normal' transmission. In his framework inspired by biology and particularly population genetics, all languages develop in a process of feature competition, selection, and then replication. There is no qualitative difference between a feature pool that comprises features from dialects of more or less the same language, and a feature pool that is composed of features from several distinct languages (cf. Mufwene 2001: 4–6). The processes of selection and stabilization of features then depends upon what Mufwene calls "language ecology". Schneider sums up the concept as follows:

> Which variants from the pool are chosen as stable elements of the newly emerging variety depends upon the complete "ecology" of the contact situation, the set of relevant conditions and circumstances, both extralinguistic and intralinguistic. All these parameters enter a complex "contact equation", which is too complex to be spelled out explicitly and allows for some degree of chance impact (ecology "rolls the dice", it is stated). They include the following components at least: the numerical (demographic) and social relationships (including mutual attitudes and power distributions) between the participants in a contact situation, the amount and types of communicative events, the nature of the linguistic input elements, surface similarities and typological degrees of relatedness between the languages involved, and so on. (Schneider 2007b: 22–23)

From this characterization of what constitutes language ecology, it is not altogether obvious how Mufwene's concept differs from Thomason's theory of contact-induced language change (cf. Thomason 2001, 2003 and Chapter 2.3 above).[13]

13. Similarities between Mufwene's and Thomasons's concepts have already been noted by Schneider (2007b: 23).

Mufwene does not spell out how the individual determinants of the overall language ecology are related to each other in shaping the new variety emerging from this ecology. The 'feature pool' appears as the largely unstructured "sum total of the individual forms and variants that each of the speakers involved, with different language backgrounds and varying linguistic experiences, brings to the contact situation" (Schneider 2007b: 22). How and why the competition of features in the feature pool should proceed from mere variation to actual selection is also left unspecified.[14]

The problematic aspects of Mufwene's concept can be illustrated with reference to a recent study of colloquial Singapore English or Singlish by Ansaldo (2009). Ansaldo proposes to account for three perspicuous features of Singlish – zero copula, predicative adjectives, and topic prominence – as "selection of non-English material from the multilingual pool of variables available in the linguistic ecology" (2009: 142). The feature pool from which Singlish emerged included, besides English, Sinitic and Malay, "the dominant languages of Singapore's linguistic ecology" (2009: 139). The feature pool for the emergence of an Asian English variety (AEV) might then look like this:

> we should not assume that English was the one and only target for non-native speakers in the evolution of AEVs; rather, the speakers that contributed to the development of AEVs were busy selecting and replicating linguistic features from a pool within which English grammar constituted but a subset of choices available. It is important to realize that normal transmission is untutored, creative and involves more than one language in most colonial settings where AEVs emerge. (2009: 138)

Such a characterization is open to criticism on several counts. First of all, it runs counter to the established definition of a New English as developing through the education system (Platt et al. 1984: 2, cf. Chapter 2.1). Platt et al.'s definition makes a categorical distinction between New Englishes and Pidgins and Creoles, based on precisely this property, and even though this distinction is no longer upheld, the transmission of English in e.g. India is generally anything but "untutored". Next, Ansaldo's characterization quoted above conjures up an image of free-floating, context-free features from several languages that are in competition, regardless of the actual communicative situation, the pragmatically dominant language in Matras's sense (cf. Chapter 2.3), or the social setting. The actual selection process for Singlish is then mostly determined by the frequency of a particular structure in the feature pool: "Numerical and typological dominance mean that Sinitic and

14. For an attempt to be more specific on the internal structure of the feature pool see Schneider (2007b: 99–112).

Malay variables are more frequent and salient and thus more readily available for selection and replication" (2009: 144).

To resort to frequency as the decisive factor for the selection of a feature out of the feature pool raises more questions than it answers. How, for example, can we account for different outcomes of language contact such as e.g. mixed languages, which combine one language's syntax with another's lexicon (cf. Matras 2000) *vis-à-vis* New Englishes, which are unquestionably varieties of English? If frequency were the crucial factor, then why has IndE not been restructured with basic SOV word order, like almost all its substrate languages that contributed to the feature pool?

To me, the value of the concept seems to be more metaphorical in nature. As already mentioned, the notion of "communicative space" developed by Koch & Oesterreicher (1996) is more useful in conceptualizing a multilingual contact scenario. A communicative space is essentially a "variational space" which includes more than one language, with "variational space" defined as "die Gesamtheit der in der Architektur einer historischen Einzelsprache gegebenen Sprachformen" (Oesterreicher 2001: 1564)".[15] The linguistic forms which make up the variational space are internally structured as well as hierarchically ordered along the parameters of regional, social, and stylistic variation.[16] Crucially, the concept then adds another dimension to the variational space, namely the parameter orality/literacy.[17] Orality and literacy should not be seen as a dichotomy, but rather as the two poles of a continuum from conceptually oral to conceptually written texts – compare, for example, the ICE-text category S2 "monologue" with its subdivision into "unscripted monologue" and "scripted monologue". A "spontaneous commentary" in the category "unscripted monologue" is more likely to be marked by such linguistic features as will be discussed in Chapter 4 than "broadcast news", which are typically written to be spoken.

One further component of Koch & Oesterreicher's model is the notion of "discourse tradition". Unlike the related concepts of "text type", "genre", or "register", the concept "discourse tradition" extends beyond the boundaries of an individual language and thus connects the variational space of a single language with the multilingual communicative space.[18] Language change within a variational

15. "the totality of the different linguistic forms comprised by the architecture of an individual historical language".

16. Oesterreicher applies Coseriu's terminology *diaphasic* (stylistic variation), *diastratic* (social variation), and *diatopic* (regional variation) (2001: 1565).

17. For details see Oesterreicher (2001: 1565–1570).

18. Schaefer (2006) shows how English reemerged as a written language in medieval England with the help of models provided by Latin and French discourse traditions.

space or contact-induced language change within a communicative space may then be described as "Prozesse des 'Wanderns' im Varietätenraum" (2001: 1565),[19] that is, a specific form originally marked as regional may become unmarked, or a linguistic feature that originated in one language may "drift" into another language that is sharing the communicative space. The important point here is that such "drifting" is far from random, but will proceed along the parameters which structure the communicative space.

To come back to the Indian constitution written in English: it is certainly legitimate to see the establishment of the British systems of law, of administration and of higher education in India as an imposition of colonial knowledge (cf. Cohn 1986). It is also certainly true that the introduction of colonial knowledge regimes changed indigenous forms of knowledge irrevocably, for better or worse. However, if we want to gain a perspective on the future of English as an Indian language, we need to narrow down our focus and consider the establishment of "English education and bureaucracy" (cf. above) in India in the 19th century as the time when English began to spread in the Indian communicative space via some highly specific discourse traditions.

Earlier, the British had already been confounded by the comparative paucity of written documents for legal and administrative purposes:

> Towards writing the Indian tradition shows a strangely complex combination of reverence and mistrust. [...I]t is not that this culture does not know writing; rather, inspite of knowing writing from a very early stage of its history it clearly uses writing quite sparingly. This peculiar configuration of knowing the gift of writing, yet abjuring its use in social transactions seems to indicate that in traditional Indian culture linguistic practice is governed by a 'theory' of distribution of functions between speaking and writing. [...] First, the intimate connection so common in European culture between institutions, i.e., the extension in the scale of social practices in time and space, and consigning things to the fixity of writing does not seem to obtain here. (Kaviraj 1992: 28f.)

Moreover, those texts that were found to codify Indian traditions were invariably composed in Sanskrit, the "language of the gods in the world of men" (Pollock 2006), as Sir William Jones famously discovered. Hindi at that time was a vernacular and mostly associated with oral discourse traditions and the elaboration and standardization of this language owed much to contact with English models (cf. Bhatia 1987, Rai 1984). That is, even on the eve of independence, the only alternative to English as the language of the constitution would have been Sanskrit, not Hindi, as Hindi had never been associated with this specific discourse tradition before.

---

19. "processes of 'drifting' within the variational space".

Further examples for the crosslinguistic importance of discourse traditions are not difficult to come by. Kachru suggests the term "Englishisation" (1994: 533) for the enormous impact that English-derived discourse traditions exerted on the Indian communicative space and notes with reference to South Asian literature: "There is general agreement that English has functioned as the main agent for releasing the South Asian languages from the rigorous constraints of the classical literary traditions" (1994: 536). This is a useful reminder that language contact is never a one-way street.

3.3.3 Outlook

The development of the Indian communicative space after Independence is a vast topic in its own right. A host of aspects are relevant here: issues of language planning and language development directed at the Indian languages (cf. Jayaram & Rajyashree 1998), language conflicts (cf. King 2008), the status of minority languages and language endangerment (cf. Bhatt & Mahboob 2008), and, last but not least, the current position of English in India's communicative space. I have already alluded to the two opposing viewpoints when it comes to the role of English in India: one camp sees English as an "outsider" and as a threat to the Indian communicative space. For some, the fact that English is not "of the land" is enough; others fear that by extending into discourse traditions formerly reserved for indigenous languages, English will threaten India's linguistic pluralism. The other camp embraces English as an asset: knowledge of English is the key to upward social mobility because it increases employment prospects dramatically. This discussion about the role of English in contemporary India has a common denominator, though:

> In the newly vibrant PCEs [postcolonial Englishes], there is less interest in attitudes towards indigenous varieties and their features, though there is of course discussion on the suitability of standard English versus locally colored usage for formal purposes like business contexts [...] What is more commonly debated is the role of English in general, both in a given society and as an international language. English has been portrayed as a means of both liberation and oppression, and the history of the language is full of strange paradoxes. (Schneider 2007b: 95)

Both the pro-English and the anti-English camp conceive of English as a largely monolithic entity, devoid of internal variation. In technical terms, the perspective on English in India adopted by both camps is that of the sociologist of language rather than that of the sociolinguist. This preoccupation is also found in scholarly contributions to the topic: a book or article with "sociolinguistics" and "India" in its title invariably discusses the position and functions of English within the overall

communicative space, rather than the IndE variational space on its own.[20] If the existence of an IndE variational space is at all acknowledged, then only with respect to the "cline of proficiency":

> The parameters determining variation [in South Asian English] include the following. The first is the users' proficiency in English in terms of language acquisition and years of instruction in the language. The second is the region of South Asia to which the user belongs and the impact of the dominant language of that region on English. [...] Thus there is a *cline of proficiency* in English. The two ends of the spectrum are marked by educated South Asian English at one end and by Broken English at the other. (Kachru 1994: 508f.)

I have argued in Chapter 2.4 above that the variational space of English in India has been changing over the last decades. Evidence is coming forward for social dialects in IndE, independent of and beyond the cline of proficiency referred to above (cf. Chand 2010). In what follows, I will provide further evidence for the internal structure of the IndE variational space.

20. Cf. Dimock et al. (1992), Agnihotri (2002).

CHAPTER 4

The syntax of spoken Indian English

4.1 Introduction

Braj Kachru stated in 1994 that describing the "major features which contribute to the distinctiveness of South Asian English" (Kachru 1994: 513) becomes a particularly challenging task in the realm of syntax:

> When we come to the grammatical characteristics of South Asian English, we are on rather difficult terrain. There is as yet no large-scale study of spoken or written South Asian English. Nor has any serious attempt been made to distinguish the features in terms of the proficiency scale, the register-specificity of the features and the distribution of grammatical features with reference to the regions.
> The available studies are either impressionistic or based on analyses of restricted texts, from which some generalisations have been made. These studies, useful as they are, leave much scope for further research. (Kachru 1994: 518)

However, this observation has not led Kachru himself to abandon his list of syntactic IndE features which follow his pronouncement (Kachru 1994: 518–521) for a more empirical approach to the study of IndE syntax. More or less the same list of features deemed somehow "typical" of IndE has been repeated since it appeared in Kachru (1986: 39f.), and in *Asian Englishes: Beyond the Canon* (2005), Kachru simply reprinted his 1994 article as Chapter 3 of the book ("South Asian schizophrenia") without bothering to acknowledge the considerable body of work on IndE syntax that had appeared in the meantime. In retrospect, Kachru's comment as published in 1994 already fails to register the existence of the Kolhapur Corpus of Indian English and the research possibilities related to it (cf. Shastri 1988). Further, the first comprehensive corpus-based study of written IndE by S. V. Parasher had appeared in 1991, well before Kachru complained about the absence of large-scale studies. Keeping in mind the release of ICE-India in 2002 and the numerous studies on IndE being published since then, the same lament about the unavailability of syntactic studies repeated eleven years later (Kachru 2005: 48) sounds grossly negligent. Over the years, Kachru's writings on IndE have remained practically untouched by current research on IndE syntax; they have established and continued the "feature-list-approach" to IndE that has been criticized by Sedlatschek (2007, 2009) as positively detrimental to further studies (cf. also Chapter 4.6.2).

The great advances in corpus linguistics over the last two decades together with a shift in perspective concerning the New Englishes (cf. Chapter 2.1) have had a considerable impact on empirical studies of IndE. Typically, earlier studies were concerned with "syntactic deviances of Indian English" (cf. Sahgal & Agnihotri 1985). Those features of IndE usage that were classified as "deviant", that is deviant from the BrE standard, were typically presented to speakers of educated IndE and/or speakers of BrE or AmE to elicit their judgments on grammaticality and acceptability. Such a "referee approach" to IndE has now largely been replaced by a descriptive quantitative-comparative paradigm. Two main research directions can be discerned here which complement each other. One the one hand, there are in-depth studies that focus on a particular New English and trace the function and distribution of selected features through different text types, depending on the size and the composition of the corpus (e.g. Balasubramanian 2009, Sedlatschek 2009 for IndE). Other studies concentrate on a particular feature and compare its function and distribution across several New Englishes, thus making full use of the available ICE-corpora (e.g. Schneider 2004, Hundt 2006, Mukherjee & Gries 2009).

4.1.1 Corpus-based approaches to IndE

Comparative studies are explicitly or implicitly driven by the question whether the New Englishes *as a class* show similarities as well as differences which set them apart from other varieties of English – be they called "native", "Inner Circle" or "L1" varieties of English (cf. Chapters 2.1 and 2.2). Such studies frequently come to two conclusions. Firstly, the frequency of occurrence of a specific feature is strongly correlated with text type, with the spoken genres much more prone to variation than the written genres. Secondly, differences between the "New" and the "Old" Englishes tend to be more quantitative than qualitative in nature. Sand, for example, summarizes her ICE-corpus study on article usage across New Englishes thus (2004: 294–5):

> In any case, the findings presented above confirm the importance of text type in linguistic analysis. In all varieties, differences across text types are observable and genre differences within one variety are practically always more pronounced than overall variation across varieties. In some text types like academic prose and instructional writing (W2A, B, D), text-type specific conventions override any variety-specific preferences. It is thus highly questionable when generalizations about the nature of 'Jamaican English' or any other variety are made on the basis of data of only one text type.

Similarly, Schneider concludes after reviewing a host of corpus-based studies of New Englishes:

> individual varieties differ from each other first and foremost in their combinatorial preferences, in their constructions, in the frequencies of their lexicogrammatical choices, collocations, word uses, and so on. It is not only, and perhaps not even primarily, the occasional occurrences of well-known "distinctive features" that attribute its uniqueness to a variety; it is the subconscious set of conventions regulating the norm level of speech habit, of what is normally said and done, the "way things are said" in a community. (2007b: 92)

Note that Schneider is implicitly appealing to pragmatics rather than syntax or the lexicon as the source for the most conspicuous distinctive features of a variety – we will return to this point repeatedly in Chapters 4 and 5.

The fact that the spoken language displays the highest degree of variation and divergence from the "Old" Englishes is not at all surprising, for several reasons. First of all, spoken language universally tolerates a much higher degree of variation than written language. Secondly, the abstract notion of "language contact" becomes very real in multilingual speakers' conversational interactions, as I have shown in Chapter 2.3 with reference to Matras & Sakel (2007). Keeping in mind what has been said about the structure of the communicative space in Chapter 3.3 above, Sand's observations about the properties of text types such as academic prose or instructional writing are also easy to explain. The discourse tradition "Academic writing in English" is sustained by a worldwide discourse community whose shared norms are more or less the same, regardless of the individual author's first language or the actual place of publication of the individual book or article. Other written genres, even though comparable in their degree of formality, may admit variety-specific usages to different degrees, depending on their degree of embeddedness in the local culture. A prominent example that comes to mind is Indian creative writing in English, a text type that provided Kachru with most of his examples for "Indianisms" in the first place (cf. Chapter 2.1.3).

Thus, Sand is certainly right in pointing out that generalizations about the "nature" of a particular variety should not be made on the basis of one single text type. However, in the following I propose to do just that: as already indicated in the Introduction, I will examine some aspects of discourse-pragmatic sentence structure in IndE exclusively on the basis of the ICE text category "direct conversations" (S1-A). Some of the preceding chapters have already touched upon the motivation for my choice (cf. Chapters 2.3; 2.4; 3.3). Before turning to my selection of linguistic features, I will briefly summarize why spoken IndE, more specifically conversational IndE, affords such a unique point of departure for the study of variation and change under contact conditions.

In Chapter 2.3, I discussed Matras & Sakel's (2007) proposal for a model of contact-induced language change. One of the crucial aspects of their model is to consider the actuation and the propagation of innovative features in a multilingual setting as two strictly separate processes: the former is mainly cognitive, the latter largely social.

> The crucial factors on the propagation of innovations in contact situations appear to be, rather, the directionality of bilingualism, and the extent of control and pressure that is exerted on speakers to conform to more established speech norms. (Matras 2009: 312)

Thus, in order to examine speakers' choices in multilingual environments, we need to focus on their interactions in conversation – in our case, the ICE-India conversation files S1A 1–100. Whether the possibilities of expression that emerge as innovations at this level are admitted into other text types is an entirely different matter. The more conceptually written the text type, the more likely it is that the codified norms of written English will not easily be compromised. Some of the features to be discussed in the following are also quite unlikely to cross the threshold to written IndE because they exemplify strategies which are generally absent from the written language. Still, I would maintain that the differences between most New Englishes and the "Old" Englishes are not only quantitative, but also qualitative, at least when it comes to their source contexts in multilingual conversation.

4.1.2 ICE and ICE-India

It is vital for any corpus-based study to address both the scope and the limitations arising from the actual corpus, its compilation and design. In the following, I will first briefly sketch the research opportunities opened up by the ICE-project in general and ICE-India in particular. Secondly, I try to assess the representativeness and reliability of the data assembled for ICE-India for the purposes of this study.

The ICE-project was initiated by Sidney Greenbaum as the first representative corpus of varieties of English around the world. The focus was explicitly placed on standard varieties of Englishes (cf. Greenbaum 1990). The initial corpus design as envisaged by Greenbaum and his associates was taken over for all subsequent ICE-corpora, and it is this structural parallelism which allows for comparative studies across a wide range of varieties as well as text types. ICE-corpora generally consist of around one million words, with the main dividing line between written (40%) and spoken (60%) texts, each individual text consisting of roughly 2,000 words.

At the time of writing, the ICE-project has published corpora for Canada, East Africa, Great Britain, Hong Kong, India, Ireland, Jamaica, New Zealand, the

Philippines, and Singapore, and many more are under preparation.[1] While all ICE-corpora are fully compatible with respect to their structure and markup, the informativeness of the individual corpus manuals varies wildly, as does the information contained in the text headers. Some corpus compilers discuss the problems that arose in the creation of their corpus in the corpus manual (e.g. Hudson-Ettle & Schmied 1999 for ICE-East Africa), thus giving a valuable assessment on the status of English within a particular country, and ultimately on the quality and comparability of the corpus data. Many, but by no means all ICE-corpora spell out sociolinguistic information about the speakers who contribute to the corpus. In ICE-India, for example, the text headers cover the independent variables familiar from sociolinguistic research, namely age, sex, education, occupation, mother tongue, and additional languages.

This study is based on the text category "private dialogue" (S1A 1–100), which comprises 100 transcripts of conversations. The first 90 of these files represent "direct conversation", the remaining ten are transcripts of telephone conversations. In sum, this ICE-India subcorpus consists of 218,531 words produced by 242 speakers, i.e. about 20% of the overall corpus.[2] The relevant subsection of ICE-GB consists of 205,627 words and samples recordings from 374 speakers.[3] Since ICE-India is a lexical corpus like all other ICE-corpora except ICE-GB, I went through all text files and added manual coding for the syntactic phenomena to be discussed in this study.[4]

For the sake of readability, all textual markup was removed from the examples, with a few exceptions: pauses (indicated by <,>) and metalinguistic comments such as [laughter] were retained.[5] No attempt was made to correct obvious spelling mistakes in the transcriptions; only where intelligibility was compromised did I add a comment.

1. Additions and extensions of the ICE-project are published on the ICE homepage: http://www.ucl.ac. uk/english-usage/ice/index.htm.

2. Information about gender and mother tongue is available for 241 speakers, information about age is provided for 237 speakers.

3. The actual number of speakers is slightly lower, since the ICECUP-statistics for ICE-GB count each speaker separately, even if he or she participates in several conversations or parts of conversations (subtexts).

4. ICE-GB is fully tagged and parsed and is distributed with the highly sophisticated search software ICECUP. For a comprehensive introduction to working with ICECUP cf. Nelson, Wallis & Aarts (2002). The manual coding was carried out with the help of the freeware editor *Programmer's Notepad* (http://www.pnotepad.org/).

5. The ICE markup manual for spoken texts is available at: http://www.ucl.ac.uk/english-usage/ice/manuals.htm.

4.1.3 The speakers

As said above, all individual ICE-corpora are representative of standard Englishes. In order to avoid any preconceived notion of "standard", the ICE-corpora include language samples from speakers who have to fulfil the following criteria to be considered "standard speakers":

> The authors and speakers of the texts are aged 18 or over, were educated through the medium of English, and were either born in the country in whose corpus they are included, or moved there at an early age and received their education through the medium of English in the country concerned. (http://www.ucl.ac.uk/english-usage/ice/design.htm)

However, such a characterization of what constitutes a speaker of a country's standard variety of English is intrinsically biased towards monolingual English speaking countries (cf. Lange 2007). For multilingual countries, several problems arise with this definition. To take one example: the compilers of ICE-East Africa have commented upon the limited role that English might play in a communicative space in which it is a second language:

> In ESL communities English is normally used in domains related to the upper part of the formal spectrum. It may therefore be difficult to find texts for the spoken private categories, because other languages are preferred in conversations among family members and friends. [...] The vast majority of the direct conversations in ICE-GB would simply not be conducted in English: all the family conversations (e.g. S1A-007) and mealtime conversations (e.g. S1A-056) in the British corpus would be too exceptional to be included in an African or Asian corpus of English. In most ESL cultures the use of English would be considered rude in such contexts, as the older members of the family might be excluded because of their lack of language skills. (Schmied 1996: 185–186)

As a consequence of the restricted range of English in East Africa, the corpus compilers did not manage to procure enough speech samples (cf. Hudson-Ettle & Schmied 1999).

Basically, a comparable division of labour holds in all countries where English as a colonial language was superimposed on an indigenous communicative space. Parasher's characterization of educated IndE echoes Schmied's comment above (1991: 51):

> The IE speech community consists of those bilinguals who use it as a second language mainly in Indian social, educational and administrative contexts. The specific functions for which English is used in India are of formal rather than informal or intimate type. [...] English is *learned* not *acquired* as L2 in a formal classroom situation from Indian teachers who themselves speak and write IE.

However, Parasher also points out that "gradually English bilingualism is increasing even in certain informal domains such as the family" (1991: 95). Written in 1991, this casual observation certainly holds true for India in the 21st century: the functional range of English is likely to increase steadily (cf. d'Souza 1997, 2001).

Another point raised by Schmied touches upon the ICE definition of "standard speaker" in ESL contexts:

> What is unique to EFL and ESL communities is the problem of learner languages, which affects the basic parameters age and education, because in a sociolinguistic context where English is learnt only as a second (or third, etc.) language it is difficult to determine where an interlanguage ends an educated English starts. ICE only collects texts from adults over 18 years of age, but the ESL corpora must be sure to include (if at all possible) only speakers who have received their formal education through the medium of English. (1996: 186–187)

For India, the requirement of English-medium education comes dangerously close to discrimination on the grounds of caste (an "atrocity" in Indian legal terminology):[6] English-medium education is not the general rule in India, but a highly valued and also highly prized commodity, affordable for only a tiny minority of the population. Even if English is introduced as a subject in grade 1 in many Indian schools now, the unequal access to English is not likely to change fundamentally in the near future. The ICE-India compilers have therefore chosen to relax the criterion of English-medium education somewhat: "The category of 'conversations' are drawn largely from the trained ELT teachers, though they have not been educated in the medium of English at all levels" (Shastri 2002). By including those IndE speakers who effectively act as the norm providers within the IndE speech community, the corpus compilers have come to a more realistic assessment of who is to be included in the corpus as representing standard IndE.

The remainder of this section will deal with the speakers' sociolinguistic profiles. Wherever possible, I will compare the relevant data of both ICE-India and ICE-GB speakers and correlate these with the respective country's population statistics.[7]

---

6. English-medium education is frequently correlated with high caste: I was told that whenever employers want to exclude dalits (i.e. "untouchables") from applying for a job, they simply add "English-medium education from primary school onwards" as a requirement in the job description.

7. All data concerning the overall Indian population come from the 2001 census which is available online: http://www.censusindia.gov.in/. Data relating to the UK is also publicly available at http://neighbourhood.statistics.gov.uk; the data are taken from the 2001 census.

4.1.3.1 *Speaker variables: Gender.* The corpus compilers were, as Table 4.1 shows, very successful in achieving a gender balance among the participating speakers; women are even slightly overrepresented. Inclusion of women may prove difficult in countries where access to education is more restricted for girls than for boys. A quick look at the all-India literacy rate according to the 2001 census confirms that unequal access to education remains a national problem:

> While the overall literacy rate works out to be 64.8%, the male literacy rate is 75.3% and that for females is 53.7%, showing a gap of 21.6 percentage points between the sexes at the national level. The gap is more in the rural areas. In the urban areas, higher literacy rate has been recorded both for males and females and the difference among the sexes is lower (13 percentage points). (http://www.censusindia.gov.in/Census_And_You/literacy_and_level_of_education.aspx)

The high representation of women in the corpus is even more remarkable when we consider the Census figures for participation in higher education: if we look only at two categories that are relevant for recruiting ICE-speakers, we find a pronounced gender imbalance: the category "Higher secondary/Intermediate Pre-University/Senior secondary" lists 24,596,339 males and 13,219,876 females, and the category "Graduate & above" includes 25,533,308 males and 12,136,839 females respectively, i.e. a rough ratio of 2 to 1.[8] Corresponding figures for England and Wales from the 2001 census are 2,545,439 males/2,238,045 females with a first degree and 1,149,478 males/989,716 females with a higher degree.[9]

Table 4.1. Percentage of male and female speakers in ICE-India and ICE-GB, compared with the proportion of men and women in the respective overall populations

	India		UK	
	In the corpus (%)	**Overall population (%)***	**In the corpus (%)**	**Overall population (%)****
male:	49.38	51.73	49.73	48.66
female:	50.62	48.27	50.27	51.33
Total	100	100	100	

* The 2001 census provides the sex ratio (females per 1000 males) for India as a whole (933), for rural (946) and for urban (900) India. (http://www.censusindia.gov.in/Census_Data_2001/India _at_glance/fsex.aspx).

** The figure covers England and Wales (Table T 07 from the 2001 census, available at http://www. statistics.gov.uk/statbase).

8. http://www.censusindia.gov.in/Census_Data_2001/Census_data_finder/C_Series/Literates_and_educational_level.htm.

9. http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=7553&More=Y.

4.1.3.2 *Speaker variables: Age.* For India, this category does not allow a direct comparison with the census data because of differing classification criteria: the age groups as specified in ICE-India and those specified by the Census do not match. What is apparent is that the age group 18–25 is overrepresented in the corpus: more than 30% of speakers in the corpus belong to that age group, compared with maximally ~18% in the Indian population, if we combine the figures for the age groups 15–19 and 20–24.

Further, data related to age are also not directly comparable between ICE-India and ICE-GB: only the youngest age group (18–25) is identical for both corpora, all other age groups are split up differently in the two corpora. Whereas the ICE-India age categories proceed in steps of seven years for age groups and do not differentiate beyond 50plus, the ICE-GB age categories change at an interval of 19 years, a rather rough classification.

4.1.3.3 *Speaker variables: Languages/mother tongues.* This category is not relevant for ICE-GB, but very important for ESL countries and indispensable for the purposes of the present study. Again, the corpus compilers have managed to include speakers from maximally diverse linguistic and ethnic backgrounds. The fact that there are such disproportionately large groups of Marathi and Kannada

Table 4.2. Speaker age in the ICE-India direct conversation files

Age groups	ICE-India (%)
18–25	30.38
26–33	23.63
34–41	18.14
42–49	9.28
50+	18.57
Total	100

Table 4.3. Speaker age in the ICE-GB direct conversation files

Age groups	ICE-GB (%)
18–25	36.08
26–45	46.59
46–65	14.2
66+	3.13
Total*	100

* Information about age is available for 352 speakers.

Table 4.4. Indian population according to age group, 2001 census (http://www.censusindia.gov.in/Census_And_You/age_structure_and_marital_status.aspx)

Age group	Indian Population
0–4	10.7
5–9	12.5
10–14	12.1
15–19	9.7
20–24	8.7
25–44	27.6
45–64	13.5
65–79	4.0
80+	0.8
Less than 18	41.1
Less than 21	47.9
Total	100

speakers is probably an accidental by-product of the sampling procedure: many conversations were recorded during an English refresher course held at the *Central Institute of Indian Languages* (CIIL) in Mysore in Kannada-speaking Karnataka. Another large set of recordings was made in the vicinity of the Shivaji University of Kolhapur (in Marathi-speaking Maharashtra), the 'home base' of the corpus compilers. However, it remains striking that the percentage of Hindi speakers represented in the corpus is just above 7%, if we include the speakers of Marwari and Bhojpuri, two "mother tongues" that are subsumed under Hindi as "language" in the Census. Speakers of Hindi form the largest speech community in India and account for 41% of the overall population according to the 2001 Census.

Table 4.5 lists all Indian languages that were returned as mother tongues by ICE-India speakers, their genetic affiliation and the number of speakers (in %), both within the corpus and across India. Table 4.6 then shows the proportional representation of the four language families of India, again as they are represented in the corpus and in India generally. Finally, Table 4.7 tries to capture patterns of multilingualism.

The table shows that speakers of Dravidian languages are somewhat over-represented in the corpus, compared to their overall proportion of the Indian population. Tibeto-Burman languages are represented by only five speakers, Austro-Asiatic languages by one speaker of Khasi. Even though Tibeto-Burman language speakers are overrepresented in the corpus in comparison to the overall Indian population, the absolute number of speakers of Tibeto-Burman and Austro-Asiatic languages is too low for valid generalizations. The same holds for the two speakers who have returned English as their mother tongue: any research project interested

Table 4.5. Mother tongues, their genetic affiliation and the percentages of their speakers in the direct conversation files of ICE-India as compared to the overall Indian population (IE: Indo-European, Dr: Dravidian, TB: Tibeto-Burman, AA: Austro-Asiatic)

	Language family	In the corpus (%)	Overall population (%)
Marathi	IE	20.79	6.99
Kannada	Dr	19.92	3.69
Tamil	Dr	12.86	5.91
Malayalam	Dr	7.88	3.21
Telugu	Dr	7.47	7.19
Punjabi	IE	6.64	2.83
Hindi	IE	5.81	41.03
Marwari	IE	0.83	0.77
Bhojpuri	IE	0.41	3.21
Konkani	IE	4.15	0.24
Bengali (Bangla)	IE	3.73	8.11
Oriya	IE	2.07	3.21
Kashmiri	IE	1.25	0.54
Urdu	IE	0.83	5.01
English	IE	0.83	0.02
Naga	TB	0.83	*
Manipuri	TB	0.83	0.14
Nepali	IE	0.41	0.28
Assamese	IE	0.41	1.28
Tulu	Dr	0.41	0.16
Sindhi	IE	0.41	0.25
Angami	TB	0.41	0.01
Khasi	AA	0.41	0.11
Gujarati	IE	0.41	4.48

*'Naga' as the designation for a language or mother tongue is not found in the 2001 Census, either because 'Naga' designates an ethnicity rather than one single language, or because it was subsumed as 'others' under some other language spoken in Nagaland. The SIL *Ethnologue* lists 40 (!) varieties of Naga: http://www.ethnologue.com/show_country.asp?name=IN.

Table 4.6. Representation of language families

Genetic affiliation of mother tongues	in the corpus (%)	all India (%)
Indo-European	48.98	76.89
Dravidian	48.54	20.83
Tibeto-Burman	2.07	1.0
Austro-Asiatic	0.41	1.11

in the future development of IndE would jump at the chance of comparing data from speakers for whom IndE is an L1 with data of IndE as L2. However, two speakers alone do not make for representative and comparable data, and a corpus of L1 IndE has yet to be created.

4.1.3.4 *Patterns of multilingualism.* The ICE-India headers also list information about speakers' other languages, a noteworthy feature in several respects in the context of this study. The degree of individual multilingualism as depicted in Table 4.7 is striking and reflects widespread societal multilingualism, a defining feature of the Indian communicative space.

Table 4.7. Speakers' multilingualism in the ICE-India conversation files

	Other languages			
Mother tongue	**None***	**Hindi**	**other(s)**	**total**
Marathi	6	44	1	51
Kannada	6	37	3	46
Tamil	14	13	3	30
Malayalam	4	11	5	20
Telugu	–	10	4	14
Punjabi	1	14	1	16
Hindi	6		9	15
Marwari	–	2		2
Bhojpuri	–	1		1
Konkani	2	6	2	10
Bengali/Bangla	3	5	1	9
Oriya	1	4		5
Kashmiri	–	3		3
Urdu	–	1		1
English	–	2		2
Naga	2			2
Manipuri	–	2		2
Nepali	–	1		1
Assamese	–	1		1
Tulu	–	1		1
Sindhi	–	1		1
Angami	1			1
Khasi	1			1
Gujarati	1			1
Sum total/in %	**48/20.3%**	**159/67.4%**	**29/13.3%**	**236/100%**

*By default, all speakers are also proficient in English, "none" applies to languages other than English.

In narrowing the presence of other languages in speakers' multilingual repertoires to "none" (except, obviously, English), "Hindi", and "others", I tried to capture the importance as well as the availability of Hindi as the national link language alongside English. The table clearly reveals Hindi to be the most pervasive second language in the ICE-India speaker sample, which is also quite likely to reflect the status of Hindi throughout the entire Indian speech community: Hindi therefore may lay some claim to be the "pragmatically dominant language" in the sense of Matras (1998, 2000) for many speakers in diverse communicative situations, and we will trace the significance of this notion in subsequent chapters. Speakers of minority languages such as Konkani, Manipuri, Sindhi, or Tulu are more likely to be proficient in both Hindi and English.

The speakers of Naga, Angami and Khasi in the corpus form an exception to this general pattern. The reason may be that both Naga and Angami are languages of Nagaland, a small northeastern Indian state which has chosen English as its official language and as the medium of education in schools. The same holds for Khasi, a language spoken in the northeastern state of Meghalaya which likewise has opted for English as official state language. Interestingly, more than half of the Tamil speakers do not speak Hindi: Hindi ceased to be taught in schools in the state of Tamil Nadu after the language riots in 1965 which formed the violent peak of a longstanding opposition to Hindi (cf. Lange 2010). Table 4.8 depicts the range of bi- and trilingualism as reported for the 1991 Census.[10] The pan-Indian patterns of multilingualism and the patterns found for the ICE-India speakers are broadly similar and raise the same questions that Jain (2003: 55) has posed:

> Konkani, Sindhi and Urdu, which are spoken in diaspora, have the largest incidence of bilingualism and, to an extent, of trilingualism. On the other hand, it needs investigation as to why mother tongue speakers of Bangla, Hindi and Tamil have shown the least inclination to learn another language. These are the three most prestigious languages and are socio-politically distinct from others. Bengalis are known for the love of their language, for Tamils it is a politically sensitive issue, and Hindi is the national language of the country.

It is interesting to see that Jain has no qualms about calling Hindi the "national language of the country", a notion that was carefully avoided in framing the relevant paragraphs in the Indian constitution (cf. Lange 2010).

However, labels aside, Hindi is undoubtedly the national link language alongside English, such that speakers with different linguistic backgrounds will resort to either one or the other in communication.

10. The category "trilingualism" was introduced for the first time in the 1991 census (Jain 2003: 55).

Table 4.8. Percentage of trilinguals/bilinguals in India according to mother tongue, 1991 census (Jain 2003: 56)

Language	tri.	bi.
Konkani	45	74
Sindhi	29	63
Panjabi	23	38
Manipuri	22	33
Nepali	20	40
Malayalam	20	29
Assamiya	13	28
Marathi	13	27
Urdu	13	38
Gujarati	12	25
Oriya	9	16
Telugu	8	21
Kannada	8	24
Bangla	5	13
Hindi	3	11
Tamil	2	19

4.1.3.5 *Speaker variables: Education.* In ICE-GB, the variable "speaker education" is binary: all speakers are assigned to either "secondary" or "university", i.e. those who have completed a university degree course. "Secondary" refers to

> the educational level of speakers and authors who have completed secondary schooling, but have not completed a tertiary (university) course. University undergraduates are therefore described as having secondary education. (ICECUP)

The category "educational.level" in the ICE-India headers, however, shows a much higher degree of internal differentiation.[11] Table 4.9 depicts an overview of the educational level pertaining to the ICE-India speakers; Table 4.10 provides the data for both ICE-India and ICE-GB speakers in a binary format ("secondary" vs. "university").

The speakers contributing to ICE-India are clearly over-educated in comparison to the speakers in ICE-GB, as Table 4.10 makes clear. They are also massively

11. In order to have comparable data to ICE-GB, and also to ensure compatibility with the search programme I used, I regularised the spelling (e.g. both spellings "master's degree" and "Master's degree" occur) and reduced the number of different designations within the category "educational.level" (e.g. designations such as "Master's degree", medicine" and "M.Litt" were subsumed under "M.A").

Table 4.9. Variable "educational level" in ICE-India S1A 1–100

all Ph.D	21.01%
all MA and M.Phil	51.74%
all other graduates	19.49%
all secondary (undergraduates and college)	7.74%
total	99.98%

Table 4.10. Comparison of speaker's educational level in the private dialogue files in ICE-India and ICE-GB[12]

Speakers' educational level	**in ICE-India**	**in ICE-GB**
Secondary	7.74%	30.09%
University degree	92,24%	69.91%

over-educated in comparison to the general Indian population with its overall literacy rate of 64.8% (cf. above). Such a bias is probably unavoidable, given the sociolinguistics of English in India.

4.1.4 The data

The linguistic community owes much to William Labov's insightful methodological considerations that arose from his sociolinguistic fieldwork. His reflections on the "observer's paradox" as well as the solutions he proposed to overcome it (cf. Labov 1972b) are now a staple textbook ingredient. However, natural spoken data in the realm of New Englishes (as envisaged e.g. by Mesthrie 2006b: 275) are not always easy to come by, as the discussion above concerning ICE-corpora in ESL countries has shown.

There are many comments in the ICE-India conversations which indicate speakers' self-consciousness:

(4.1.1) B: what you are recording this for <,,>
A: Uhm <,> oh <,> it's a project <,> called the I C E <,,>
B: I C who <,> [laughter]
A: I C E <,> Interational Corpus of English <,,> We have to record <,> samples of <,> English from all over India <,,> And later on these will be analysed <,,> So we have to collect them first and that's why recording

12. Information about educational level in ICE-GB was only available for 329 speakers (99 speakers with secondary, 230 with university education).

B: In other words I'm a guinea pig <,> [laughter]
A: Well uhm <,> yes <,> [laughter] if you like to call yourself a guinea pig then yes <„> (ICE-IND: S1A-055#37–45)

(4.1.2) Just don't laugh *yaar* so much <,> because that will be recorded <,> (ICE-IND: S1A-053#11:1:A)

Two speakers even complain about the difficulty in sustaining a conversation in English:

(4.1.3) B: But we are not used to speak English <,> it's a <,> it's very difficult to speak twenty minutes English is very difficult
A: Uhn <,>
B: In Marathi we can do it know <,> or in in your Kannada <,> uhn earlier (ICE-IND: S1A-086#87–89)

This self-consciousness about being recorded, however, is not restricted to the speakers participating in ICE-India. Similar comments can be found in the ICE-GB conversations, e.g.:

(4.1.4) I do apologise for that
I'm sorry
Can we can we scrap that [unclear-word]
[laugh]
Edit that one out <„> (ICE-GB: S1A-006#265–269)[13]

(4.1.5) God I hope Laura doesn't play this tape to anybody (ICE-GB: S1A#042–315)

In order to arrive at more natural spoken data despite the overt recording, the ICE-GB compilers decided

> not to use the first fifteen minutes of recorded conversations. 2,000 words corresponds to about fifteen minutes of conversation, so in most cases at least twice this amount was actually recorded. (Nelson 1996: 31)

In at least one of the ICE-GB conversations, this fact is explicitly mentioned:

(4.1.6) A: Have you still got that damn machine on
B: Uhm yes <,>
But you can't you see you're not meant to notice it
It's got to be the tape's got to be on for
They disregard the first fifteen minutes anyway <,> because they reckon that it won't be natural

13. In this conversation, the individual turns are not assigned to speakers.

A: Oh but that's only if the people know it's on uh <,>
B: Yeah (ICE-GB: S1A-047#72–78)

Most ICE-GB conversation files indeed are excerpts from conversational exchanges that have been going on for a while, unlike the ICE-India conversations, which typically begin with a formal round of introduction, sometimes even with a brief discussion about what the topic of the conversation should be. A typical case is the following:

(4.1.7) A: I'm Mrs Gayatri Murthy <,>
I'm from Banglore <„>
B: I'm Miss Vidya Shetty
I'm from Manglore <,>
I teach in St Andrews College <„>
I'm a lecturer in English <„> (ICE-IND: S1A-031#1–6)

Sometimes speakers seem to have difficulties in keeping a conversation going or finding the right topic to begin with, as the examples illustrate:

(4.1.8) What we'll talk about <„> (ICE-IND: S1A-021#10:1:A) (uttered at the start of a conversation after the speakers introduced themselves to each other)

(4.1.9) We have ten minutes more to go so you better think of something to talk about <„> (ICE-IND: S1A-055#199:1:A)

(4.1.10) B: So I think it's already <,> two <.> min </.> it's already ten minutes more than ten minutes *no*
A: Let's hope <,> another another few lines and this
B: Okay
A: It will be over <,> so I remember <„> [...]
(ICE-IND: S1A-028#102–105)

(4.1.11) A: So total eight minutes are over we need to continue *na*
B: Uhm
A: We need to continue *na*
B: Yeah <,> we must continue other ten minutes [laughter]
(ICE- IND: S1A-045#161–164)

Again, finding something to talk about unselfconsciously does also not come naturally to all ICE-GB speakers, as the following exchange indicates:

(4.1.12) A: How much longer do we seriously have to talk for <,>
C: Well I presume it's until the tape runs out
A: No
It's only for fifteen minutes

C: Well there's only about fifteen minutes left on the tape
A: I'm trying so hard to concentrate on this
(ICE-GB: S1A-038#268–273)

So, while some of the ICE-India conversations may appear slightly unnatural and/or formal, the data are the best we have for the time being, and fully compatible with those collected for ICE-GB and other ICE-corpora. The sense of self-consciousness which is evident in the direct conversation files seems unavoidable, given the rather rigorous ICE-corpus design with its main aim of assuring representativeness of speakers and text types. A different approach to corpus compilation such as the one adopted for the *Corpus of Colloquial Singapore English* (cf. Lim 2004) may yield more spontaneous speech data, but then has to sacrifice the requirement of representativeness: Lim and her associates equipped some Singaporean undergraduates with recording devices over an extended period of time, letting them decide when and what to record. The data produced by these speakers are surely more natural, as participants had more control over the recordings and more time to get used to the presence of the recording device. On the other hand, a balanced sample of the speech community which allows for quantitative sociolinguistic studies is not attainable with this method of data collection.

The manual accompanying ICE-India closes with the somewhat cryptic comment:

> Those who will be using our corpus, it is necessary to highlight the country in which it was assembled. India is not just a country but it is a sub-continent. It is both multilingual and multicultural society. It is 'unity in diversity' as it was put by Jawaharlal Nehru. Those who have actually seen India that is Bharat in close quarters should judge this corpus for what it is worth. (Shastri 2002: 3)

In the following, I will show how "multilingual and multicultural" speakers of IndE draw upon their composite linguistic repertoires to negotiate communicative interactions. As will become increasingly clear throughout this chapter, there is indeed evidence for pan-Indian "unity in diversity" when it comes to the discourse-pragmatic features to be investigated.

4.1.5 Discourse-pragmatic sentence structure

"Information structure" has come to designate a perspective on grammar which focuses on speakers' manipulations of basic (or canonical) word order of their respective languages for specific discourse effects, specifically for the marking of topic and focus of a proposition. Such non-canonical sentence types may either simply involve a deviation from the basic linear order or further elaboration of the canonical sentence structure. In Huddleston & Pullum's terms, speakers may resort

either to "reordering" or "realignment of semantic and syntactic elements", where "realignment is accompanied by the addition of one or more elements" and therefore leads to "greater syntactic complexity" (2002: 1367). Ward & Birner restrict their discussion of information structure to constructions that may be subsumed under 'reordering', i.e. to "non-canonical English constructions in which some constituent is placed in a non-canonical position, leaving its canonical position either empty or filled by a non-referential element" (2004: 173). Relevant constructions would then be topicalization and existential *there*-constructions, whereas left- and right-dislocation as well as cleft constructions would be excluded: "such constructions place very different constraints on their non-canonically positioned constituents" (*ibid.*). Their generalization that "the non-canonical construction provides speakers with a way of ensuring that old information precede new" (*ibid.*) allegedly does not hold for the syntactically more complex realigned constructions. However, I would argue that this greater syntactic complexity seems to lie in the eye of the (linguistic) beholder. A left-dislocation construction (LD) is superficially similar to a topicalization construction in that an argument of the verb is placed in sentence-initial position. The greater syntactic complexity derives from the fact that in the LD construction, a referential resumptive pronoun is introduced:

(4.1.13) I <,> totally disagree that the women **they** are not taking up any initiatives <,,>
See for the last two or three decades as <,> or so <,> we find that women <,> **they've** reached <,,> a real good status to over the men <,> in each and every field that they have participated <,,>
Say now <,> we find that that women **they** are <,> given a voice <,,>
(ICE-IND: S1A-011#44–46)[14]

Dislocation constructions have given sleepless nights to syntacticians who try to account for their underlying structure in terms of a purely configurational syntax, this also holds true for cleft constructions and the precise syntactic status of the 'relative-like' clause, e.g.

(4.1.14) It was you that told me that wasn't it (ICE-GB: S1A-099#272:2:A)

Clefts such as the example above are indeed syntactically more complex than their canonical counterparts in that they impose a biclausal syntax on a single proposition (cf. Lambrecht 2001b: 466), and they appear more complex than dislocation constructions which still preserve the proposition within one clause, albeit with an extra-posed constituent. I will return to clefts in Chapter 4.5 below.

14. Incidentally, the speaker is male.

Coming back to the notion of "syntactic complexity": I would argue that this complexity is irrelevant for the speaker and the listener, on the contrary: if we take *discourse* rather than *sentence* as the minimal unit against which judgments of "complexity" can be made, then dislocation constructions are actually less complex (cf. Chapter 4.4. below). If we want to obtain an overview of the domain of topic and focus constructions in IndE, then it would be myopic to exclude constructions such as left diclocation because of some artificial notion of syntactic complexity. The resumptive pronoun in an LD construction has also been analyzed as a topic marker, that is, an LD construction can also be understood as a way of marking the topic not syntactically, but lexically. Rather than re-arranging the word order to foreground the topic, the speaker adds a lexeme which marks the topic, namely the resumptive pronoun.

In the following, I will discuss some aspects of information structure that have been noted in various New Englishes, together with some that are unique to IndE. The domain of syntactic topic and focus constructions includes the existential construction, topicalization, dislocation, and cleft constructions. While the existential has a specifically Indian realization that will be discussed in Chapter 4.2, both topicalization and dislocation have been observed to occur with a high frequency in a range of New Englishes (cf. Mesthrie 2008: 633). Cleft constructions have so far hardly been discussed with reference to the New Englishes, and we will see in Chapter 4.5 that it is rather their absence from spoken IndE that has to be explained.

As already mentioned, languages provide different strategies for marking information structure: syntactic means such as reordering or realignment of sentence patterns can be combined with shifts in prosodic prominence. Unlike many Indian languages (cf. Chapter 3.2.3), English does not have topic-marking lexemes or bound morphemes, but it has a wide range of focus markers (cf. Chapter 4.6.1). Thus, whether by intonation, manipulation of word order or by marking constituents with particles, speakers are able to foreground or background the relevant referents, structuring information according to their point of view. Marking some constituent (and in turn its referent) as topic or focus imposes a structure on a proposition which reflects the speaker's assessment of the prominence of a particular referent within the ongoing conversational interaction. However, the marking of topic and focus is just one aspect of the overall realm of discourse organization. Kortmann & Szmrecsanyi, for example, include lack of inversion in questions, "*like* as a focussing device (e.g. *How did you get away with that like*?)" (2004: 1148) and quotative *like* (e.g. *And she was like "What do you mean?"*) in their category "Discourse organization and word order".

I propose to delineate the notion of 'discourse organization' somewhat more specifically with reference to speaker's discourse management. For example, lack of inversion in questions is surely a syntactic feature in IndE that is worthy of

attention, but the construction has no discourse-pragmatic relevance: the presence or absence of inversion in questions does not correspond to a difference in a speaker's assessment of a proposition.[15] Other linguistic devices which speakers have at their disposal to structure their conversational interaction are discourse markers or pragmatic markers.[16] Brinton's definition of "pragmatic marker" spells out how the function of such items is related to syntactic strategies of discourse organization. The definition emphasizes both the "textual" and "interpersonal" functions served by discourse markers:

> In function, pragmatic markers serve a "textual function" of language, which relates to the structuring of discourse as text, and/or an "interpersonal function" [...], which relates to the expression of speaker attitude and to the organization of the social exchange. Among the textual functions are those of claiming the attention of the hearer, initiating and ending discourse, sustaining discourse, marking boundaries, including topic shifts and episode boundaries, constraining the relevance of adjoining clauses, and repairing discourse. Among the interpersonal functions are subjective functions such as expressing responses, reactions, attitudes, understanding, tentativeness, or continued attention, as well as interactive functions such as expressing intimacy, cooperation, shared knowledge, deference, or face-saving (politeness). Together, the textual and interpersonal functions constitute "pragmatic" meaning. (Brinton 2006: 310)

The textual function of discourse markers then overlaps with the functions served by the syntax of information structure. From the perspective of the speaker who is actively managing his/her participation in an ongoing conversation, the two realms of syntax and pragmatics are probably not discrete, but poles on a continuum of possibilities for discourse management. The remainder of this chapter will then consider several of the available devices for discourse management, with a perspective on the similarities and differences in speaker's choices within a unified domain.

4.2 Non-initial existential *there*

The IndE construction *library facility was not there* (ICE-IND: S1A-064#160:1:A) has been labelled "non-initial existential *there*" (Balasubramanian 2009: 111), "use of adverbial *there* for 'dummy' *there*" (Trudgill and Hannah 2002: 132) or "[u]se of post-verbal adverbial *there* in place of dummy *there*", one of the "miscellaneous features" of IndE syntax (Bhatt 2008: 560). Curiously, this feature of IndE syntax

15. Sedlatschek (2009: 289–299) found lack of inversion to be rather rare in his corpus of IndE.

16. Cf. Lange (2009) for a comparison of the relevant terminology.

has been shunned by researchers, despite the fact that it occurs much more frequently than, say, the much-quoted invariant tag (cf. Chapter 4.4.1). The only study where this construction receives more than a passing remark is Balasubramanian's corpus-based account (2009); however, this study does no more than provide figures for the register distribution of selected syntactic phenomena. Balasubramanian's figures show that non-initial existential *there* firmly belongs to the realm of spoken IndE: 18.5% of all existential constructions in the spoken part of her corpus and 1.5% in her written corpus are of the "Indian" type (overall: 5.9%) (Balasubramanian 2009: 112). So far, this construction is not attested beyond the Indian English speech community: it is not found in either SAIE or other New Englishes. This limited distribution makes non-initial existential *there* a prime suspect for substrate influence, however broadly conceived. The following discussion will again show that "substrate influence" is only the beginning, never the end of a satisfactory analysis. I will first consider briefly the discourse functions of existentials in English generally. In a next step, I will try to determine whether the IndE non-initial existential is really just the mirror image of the existential, with a change in word order but preserving the same discourse functions. These considerations will lead straight into a detailed investigation of the data from the ICE-India conversations.

4.2.1 Definition

As is only to be expected, scholars of information structure differ as to the details of their treatment of the existential construction: Ward & Birner, for example, subsume the existential under their 'postposing' category (2004: 163), whereas Huddleston & Pullum differentiate between postposing and the existential (2002: 1363–1396). Lambrecht, on the other hand, considers the existential construction and the deictic construction ("*Look, there's John!*") as sub-types of the presentational construction (1994: 179): the deictic construction refers to the text-external world, whereas the existential refers to the text-internal world. Whatever the subtleties of analysis, these and other accounts are unanimous when it comes to the main discourse-pragmatic function of the construction under scrutiny: the existential construction "serve[s] to introduce previously unidentifiable or inactive referents into a discourse" (Lambrecht 1994: 179); it "is characteristically used to introduce addressee-new entities into the discourse, and for this reason the displaced subject NP is usually indefinite" (Huddleston & Pullum 2002: 1396).[17] Note that "addressee-new" or "hearer-new" imply "discourse-new", but not vice versa, as Table 4.11 clarifies.

17. For a thorough discussion of cases where the postposed subject NP is definite see Birner & Ward 1998: 113–141.

Table 4.11. Parameters of information status (Ward & Birner 2004: 156, quotes from Birner & Ward 1998: 15)

	Hearer-old	**Hearer-new**
Discourse-old	**Evoked** "Information which has previously been evoked in the current discourse, and which the speaker therefore believes is known to the hearer."	(non-occurring)
Discourse-new	**Unused** "Information which has not been evoked in the current discourse, but which the speaker nonetheless believes is known to the hearer."	**Brand-new** "Information which has not been evoked in the current discourse, and which the speaker does not believe to be known to the hearer."

The following conversational exchange from ICE-GB illustrates this function of bringing hearer-new referents into the discourse:[18]

(4.2.1) A: Was it good
B: Yeah
S set in working class Dublin <,> Unknown cast <,> **There's** a guy who lives with his family and <,> his father's into Elvis <,> and this
A: Mm
Yeah
Mm mm
So what sort o what sort of music do they play
B: Sixties soul <,>
A: Uh Otis Redding stuff
B: He wants to get a band together <,> so he asks all these well asks around advertises in the paper and gets all these weirdos coming to the door <,> like punks and heavy metal and <,> Trying to get a band together **There's** about <,> three girl singers <,> **There's** a soloist and drummer And they all get together but they all have fights and It's a very good film though (ICE-GB: S1A-025#1–19)

The first existential construction in B's turn, *There's a guy who lives with his family*, introduces the main protagonist of the movie she is summarizing for A. The "guy" has not previously been mentioned and represents information that is new in the

18. Both speakers are 26–45 yeas old, speaker A is male, speaker B is female.

discourse and new to A. *There* as dummy subject occupies the topic position which is generally associated with old information, whereas the actual new information, *guy*, appears in the postverbal comment position. In Ward & Birner's terms:

> postposing constructions preserve the old-before-new information-structure paradigm 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. (2004: 163)

Clause-initial *there* clearly "has the status of an empty grammatical element" (Biber et al. 1999: 1391), comparable to dummy *it* in extraposition constructions. The form has lost any traces of a locative meaning and is generally unstressed. Further support for considering *there* the grammatical subject of the clause comes from the fact that it appears in tag questions, e.g.:

(4.2.2) You know there are few people like that in whatever field <,> aren't **there**
(ICE-GB: S1A-045#008:1:B)

The postposed constituent *a guy*, on the other hand, is the "notional subject" or "displaced subject" of the clause. B continually resorts to further existential constructions for the introduction of new referents into her narrative of the movie. The second existential, ***There's*** *about <,> three girl singers*, displays the lack of agreement between the notional postverbal subject in the plural and the copula, showing that the existential construction is already largely grammaticalized (cf. Biber et al. 1999: 94). This existential construction and the one to follow, ***There's*** *a soloist and drummer*, differ from the first in that the topic expression is not modified by a following relative clause. Huddleston & Pullum (2002: 1393–1396) make a distinction between such "bare" and "extended" existentials, where "extension" refers not only to relative and non-finite clauses, but also to locative or temporal complements, e.g. *There's a friend of yours at the door* (2002: 1394).

Lambrecht's notion of the extended existential is somewhat narrower. He applies the label "bi-clausal presentation construction" to complex existentials in which the notional subject is modified by a relative clause or a non-finite clause which is frequently participial (Lambrecht 1994: 180), e.g.:

(4.2.3) **There's** an enormous racket **going on** with Edam cheese
(ICE-GB: S1A-061#304:1:B)

(4.2.4) I mean I think **there's** two people **involved** in it
(ICE-GB: S1A-066#102:1:B)

In my data, bare existentials are much more frequent in both BrE and IndE (cf. Table 4.12).[19]

Such "bare" existentials nevertheless do more than to assert the mere existence of an entity, as Lambrecht has pointed out:

> From the point of view of information-structure analysis, the label "existential" is somewhat misleading. Mere assertion of the existence of some entity is a rather special kind of speech act which is of limited use in everyday communication. [...] From the discourse-pragmatic point of view, it is therefore preferable to interpret the function of such sentences as that of presenting or introducing a referent into the "place" or "scene" of the discourse and thereby of raising it into the addressee's consciousness, rather than of asserting its mere existence. (1994: 179)

In fact, even though existential constructions may be employed to predicate the existence or non-existence of some entity (e.g. "There are no hobbits"), it is obvious that existential predication is not confined to the existential construction and can be achieved by other means (e.g. "Hobbits don't exist"). Further, bare existentials do not have a non-existential counterpart, which indicates that the predication of existence is not their primary function (cf. Huddleston & Pullum 2002: 1393): the existential *There's a soloist* in example (4.2.1) above cannot be rephrased

Table 4.12. Bare vs. bi-clausal existentials in the direct conversation files of ICE-GB and ICE-India

	Bare existentials	Bi-clausal existentials	Total	per 10,000 words
Canonical existentials in ICE-GB	570 (76.3%)	177 (23.7%)	747	36.32
Canonical existentials in ICE-India	594 (83.2%)	120 (16.8%)	714	32.67
Non-initial existentials in ICE-India	311 (99%)	3 (1%)	314	14.37

19. The only corpus-based study of the form and function of existentials that I am aware of is Martínez Insua (2004), who finds that extended existentials in the sense of Huddleston & Pullum (2002) are much more frequent than bare existentials both in spoken and in written language. Her figures, however, should be taken with more than a grain of salt. First of all, her data are based on her own subcorpus of the *British National Corpus*; and except for stating that her subcorpus comprises about 1mio. words (50% spoken, 50% written) and reflects the overall structure of the BNC, no further information about the actual texts making up the subcorpus is given. Next, the overall number of existentials retrieved from her corpus seems to be 2690 (Martínez Insua 2004: 78), this figure mysteriously dwindles to 2410 in all further statistics. There is clearly a need for further reliable corpus-based research in this area.

as **A soloist is*, which again lends support to Lambrecht's characterization of the functional range of existentials.

This discourse function of establishing a new topic in the discourse is also apparent in the following example from ICE-India:

(4.2.5) B: And then in under this graduate classes <,,> uh <,,> something very interesting happened
The university introduced <,> courses on language <,>
A: Ahn <,>
B: And **there was a professor** <,> I won't name <,> who was trained in CIEFL <,> in the earliest CIE and then in <,,> uh Pune <,> in linguistics <,>
A: Ah ahn
B: But when he was asked to teach that course <,> he refused <,>
(ICE-IND: S1A-082#10–16)[20]

Speaker B, a linguist, is telling A, another linguist, how he became interested in linguistics in the first place. After setting the scene with a description of his undergraduate years, B introduces a new protagonist, *a professor*, who becomes the topic for B's further narrative: "Thus, it is newness with respect to the hearer's knowledge that is required for the felicitous use of existential *there*-sentences" (Birner & Ward 1998: 106). I will come back to the canonical existentials in ICE-India below.

If we now turn to the non-initial existential *there*-constructions of IndE, we find that the actual construction clashes with Ward & Birner's account:

(4.2.6) A: What happen you know I forget to tell [laughs] that guys <,> and I didn't tell you that I will tell uhn <,> What happened you know that fellow **is there** *na* I told you that he'll be he will be staring at me in the college I wanted to show you that fellow <,,>
(ICE-IND: S1A- 052#258–260)

The conversational settings in the ICE-GB example above and in this excerpt from an ICE-India conversation are similar: in both cases, the speakers have embarked on a new topic and are now proceeding to give a lengthier account of their experience – either an imaginary one as in the case of the ICE-GB movie narrative or a real experience. When A introduces a new referent into the discourse, namely *that fellow*, with the non-initial existential construction *that fellow* ***is there*** *na*, the

20. CIEFL: *Central Institute for the Study of English and Foreign Languages*, Hyderabad. CIEFL is generally regarded as the best location throughout India for both study and research concerning the English language (now renamed EFLU: *English and Foreign Languages University*).

discourse-pragmatic function ascribed to the general English existential is effectively lost. The new referent remains in subject position and therefore in the position canonically ascribed to topics, i.e. old information, whereas the dummy element *there* ends up in the preferred position for new information. In terms of information structure, the non-initial existential *there*-construction is then a rather paradoxical construction. The label "non-initial existential *there*" suggests that the construction is just the mirror image of the general English existential. However, reversing the order of a construction has obvious consequences for its information structure, and the implied parallelism between the two "existentials" might simply not exist. Further, even a cursory glance at the ICE-India data turns up many examples which blatantly violate the pragmatic constraint on felicitous "canonical" existentials, namely that the "notional" subject represent hearer-new information.[21] In example (4.2.7), two students attending a college of engineering are talking about their courses. A answers B's question about a particular course first with *Yeah it was there* and then repeats her answer by combining the non-initial existential construction with left dislocation, so that the pronominal "notional subject" of the existential is repeated as the resumptive pronoun of the left dislocation (LD) construction.[22] Anaphoric pronouns by definition do not introduce hearer-new entities into the discourse, but establish discourse-old links to the preceding context.

(4.2.7) B: Do you have this structure subjects with it
A: I have to go for Structures <,>
B: I mean the strength of material
A: Yeah **it was there**
Strength of material **it was there** <,,> in the second semester
(ICE-IND: S1A-061#115–120)

Example (4.2.8) similarly features a discourse referent as "notional subject" that has just been previously mentioned.

(4.2.8) B: Yeah eatables I think <,,> *samosas* will do <,,>
A: *Samosas* yeah that's not a bad idea then the chips
B: Because they are chi uh cheap also <,>
A: Yeah **chips can be there** <,> (ICE-IND: S1A-062#85–88)

21. Obviously, the "notional" subject is also the syntactic subject of the IndE non-initial existential; I will, however, continue to use the label "notional" in inverted commas to indicate the focus on the constituent's information status.

22. Left dislocation will be treated in more detail in Chapter 4.4.

Examples such as these cannot be reconciled with the requirements for felicitous postposing: "Thus, whenever an NP represents a hearer-old entity, regardless of its discourse status, it may not be felicitously postposed in an existential *there*-sentence" (Birner & Ward 1998: 103).

The task of this chapter is then to give a detailed analysis of the meaning and distribution of the two existentials in spoken IndE. Another example from ICE-India shows that both constructions may occur in the same conversation and even in the same speaker's turn:

(4.2.9) A: Those terrorists they have to leave <,> If uh <,>
B: But there some tourist is **there** or no <,> in Kashmir <,>
A: **There** are no tourists Now the tourism is at nil <,,> **There** are so many hill stations and uh <,> beautiful lakes and forest and everything <,> but all in vain now <,> Only the export will be **there** <,> because the fruits **there** are lot of apples and peaches and plums and pears <,> (ICE-IND: S1A-054#49–55)

In what follows, I will first describe my retrieval strategy for coding the general and the non-initial existential *there*-construction in ICE-India. Even though the examples above have demonstrated that the IndE non-initial existential might have quite different discourse functions, I will proceed with a weaker version of the generally held opinion and assume that the two constructions are at least comparable, if not equivalent. I will then consider the form and function of the non-initial existential construction in more detail.

4.2.2 Existentials: Corpus evidence

This section is concerned only with existential *there*, that is, constructions with dummy *there* as subject followed by the verb *be*. Presentational *there*-constructions as in *There remain, however, three problems with the grammaticalisation model* (Matras and Sakel 2007: 840) are not considered.[23] For ICE-GB, existential constructions are easy to retrieve: all instances of existential *there* in ICE-GB are tagged EXTHERE, which allows a straightforward search for relevant examples. There are 747 tokens of existential *there* (tagged) in 707 text units in the ICE-GB conversation files. Sixteen of these occur in tags, as in example (4.2.2) above and as in example (4.2.10):

(4.2.10) but there's not much point in that **is there** [laugh] <,,> (ICE-GB: S1A-038#223:1:A)

23. For a characterization of presentational *there*-constructions see Birner & Ward (1998: 106–113).

The string *is there,* which is so frequent in ICE-India, occurs only as part of a question or as tag.

In the manual tagging of the ICE-India conversation files, I looked for both clause-initial and clause-final *there*-tokens in copular clauses. Tagging the "canonical" existential was straightforward and turned up 714 examples, a figure comparable to ICE-GB. However, a closer look at the ICE-India conversation files revealed that not all instances of clause-final *there* automatically qualified as non-initial existential. Constructions with clause-final *there* may be locative/deictic, e.g.:

(4.2.11) And Dr Kasiwal was also **there** <,> And he taught us <,> phonetics that time <,,> and I know him since then <,> (ICE-IND: S1A-013#218–219)

(4.2.12) B: And wasn't Gokhale **there** <,>
C: Oh he was **there** (ICE-IND: S1A-091#248:–249)

In both examples, the speakers talk about individuals that were present at a particular place or occasion. Such examples are reminiscent of Huddleston & Pullum's hint at a diachronic explanation for the origin of existential *there*:

> Historically, dummy *there* derives from the locative *there* of, for example, *Don't leave your shoes there.* Locative *there* is an intransitive preposition contrasting with *here*: it has deictic and anaphoric uses [...]. *there* has been bleached of its locative meaning and reanalysed as a pronoun. (2002: 1391)

Studies that are concerned with the historical development of the existential are few. Breivik (2005) assumes that both meanings of *there*, the locative-adverbial and the dummy subject-nonreferential, were already available in the early Germanic languages. He finds examples for both uses in the earliest Old English records. He further claims, somewhat paradoxically, that "the Old English topic-marker $there_1$ was syntactically reanalysed as an empty subject when English changed typologically from verb-second (TVX) to verb-medial (SVX)" (Breivik 2005: 167), but that "the signal function of $there_1$ has evolved during the Modern English period" (2005: 175), as *there* as dummy subject was still optional in Early Modern English.

Even though the trajectory from locative to existential *there* in English still seems somewhat elusive, it is reasonable to assume that there is a cline from deictic/locative meanings of *there* to more abstract meanings.[24] In the manual tagging of ICE-India, constructions such as the examples (4.2.11) and (4.2.12) above

24. Lambrecht comments on the "formal similarity between the existential construction and the deictic presentational construction", concluding that "the presentational, location oriented function of the construction is in fact the fundamental communicative function of existential sentences" (1994: 179).

that involve proper nouns of human referents where *there* was clearly a locative adverbial were excluded.

Clause-final adverbial *there* further features in combination with an existential construction in both ICE-GB and ICE-India, e.g.:

(4.2.13) **There's** the plot of a novel **there** <,> (ICE-GB: S1A-020#142:1:C)

(4.2.14) And **there's** just the two of them and they have to stop because **there's** these two highway robbers **there** <,> thieves whatever (ICE-GB: S1A-041#310:1:A)

(4.2.15) **There** is a little boy **there** <,> uhm <,> so my son goes and plays with that boy <,> (ICE-IND: S1A-030#280)

(4.2.16) And the situation in Kerala <,> **There** are <,> many teachers **there** (ICE-IND: S1A-035#159–160)

(4.2.17) But **there** is some other implication **there** <,> which might increase the cost <,> that is what I am worried (ICE-IND: S1A-094#38:1:A)

Again, such instances of final adverbial *there* were not considered. The situation is different with another kind of 'double strategy', involving both initial and final *there:*

(4.2.18) I'm sorry <,> uh <,> teaching in the college is better than <,> uhm <,> the school
Because in school **there's** lot of responsibilites are **there** (ICE-IND: S1A-065#301–302)

(4.2.19) Now there is <,> school also **there** is some fancy dress competition is **there** on Saturday <,,> in Shahu (ICE-IND: S1A-086#38)

(4.2.20) Then <,> we have to means <,> uh that prepare some heavy <.> fo <,> food means *pau bhaji* <,> ah <,> and all that <,> uh *bissi* and <,> twenty *rupees bissi* <,> and **there** are fifteen <,,> ladies are **there** (ICE-IND: S1A-086#250)[25]

Since final *there* is in construction with a copula in these examples, it was counted as an instance of the "Indian" existential, to use Balasubramanian's term. Table 4.13 summarizes the figures.

Recall that Balasubramanian (2009: 112) found 18.5% of all existentials in her spoken corpus to be non-initial existentials. Her spoken corpus also contains scripted material such as radio shows and public dialogue such as interviews. The higher percentage of non-initial existentials in the ICE-India conversation files might then be due to the higher degree of conceptual orality of direct conversation. Non-initial

25. *Pau bhaji:* an Indian snack (spicy vegetable pulp eaten with roasted bread rolls).

Table 4.13. Frequency of existential constructions in the ICE-India and ICE-GB conversation files

	ICE-India	ICE-GB
All existential constructions	1028 (100%)	747 (100%)
Initial *there*	714 (69.45%)	747 (100%)
Non-initial *there*	314 (30.54%)	–

existential *there*-constructions would then be firmly placed in the vernacular, in spontaneous discourse. I will return to the significance of this point below.

Both existential constructions are quite common in the ICE-India conversations, and both are fairly evenly distributed: "canonical" existentials occur in 95 texts, non-initial existentials in 82 texts. Further, the constructions are not in complementary distribution: only three texts have non-initial existentials to the exclusion of the "canonical" existential, and 16 texts have "canonical" existentials, but no non-initial *there*-constructions – which leaves 79 conversations which feature both constructions. Only two conversations are devoid of either existentials. Figures such as these suggest that both constructions are firmly entrenched in the pan-Indian speech community. The question whether there are patterns of variation according to speakers' age, gender, education, or mother tongue, will be taken up below. But before I embark on any further quantitative and qualitative comparisons, it is necessary to establish the discourse properties of the IndE non-initial existential *there*-construction.

4.2.3 Non-initial existential *there* in ICE-India

I have already indicated above that many examples of non-initial existential *there*-constructions in my data fail to conform to the basic criterion for a felicitous existential, namely to introduce new referents into the ongoing discourse (cf. examples (4.2.7) and (4.2.8) above). In other words, whereas the information status of the "notional" subject has to be hearer-new for a canonical existential, the overwhelming majority of all "notional" subjects in the non-initial IndE construction represent hearer-old referents. A substantial proportion of these are also discourse-old. I will turn to these cases first.

4.2.3.1 *Discourse-old "notional" subjects.* Out of a total number of 314 examples of non-initial existential *there*, 136 examples (43.3%) feature "notional" subjects that are discourse-old. Among these, several groups may be distinguished: first of all, there are those examples where the "notional" subject is discourse-old by

definition, that is by virtue of being realized by a discourse-deictic or anaphoric pronoun which establishes an explicit link to the prior discourse. The pattern 'anaphoric pronoun + be + *there*' is represented by 27 tokens, and the collocation *that is (also) (not) there* by 24 tokens:

(4.2.21) A: It depends everything on the headmaster you know
If the head is good everything is good
B: Yeah **that is there** (ICE-IND: S1A-065#274–276)

(4.2.22) B: But these food habits they differ from country to country
A: Yeah
B: Generation to generation <,> state to state <,> family even the individual <,,> each individual
A: It depends it depends of on one's likings and dislikings
That is also there <,> and it may be due to climate also
(ICE-IND: S1A-072#15–19)

(4.2.23) B: This young generation the young persons <,> they are so used to this <,> or may be they don't have so much time at their disposal now <,> they want readymade food <,> so they go for junk food
A: *Haan*[26]
Oh <,> readymade
B: The fast food
A: *Haan* <,>
B: The canned food <,> all that <,> don't you think so <,>
A: Yes **it is there** but uh <,> directly they will be affected by
B: *Haan* **that's there** (ICE-IND: S1A-072#44–51)

(4.2.24) A: Uh yesterday I just wanted to ask you one thing uh
B: What about?
A: The jeweller <,> that you go to <,> because I wanted to take a new *mangalsutra* from for me <,,> and[27]
B: Okay okay <,,> [laughs]
A: I went there <,> just went there and uh <,>
B: Where did you go <,,>?
A: Uh <,>
B: To the city
A: Uh yeah **it's there** in the city in the evening I went there <,> to jeweller shop <,,> (ICE-IND: S1A-075#1–9)

26. *Haan*: Hindi 'yes'

27. *Mangalsutra*: a necklace; symbol of Hindu marriage (comparable to a wedding ring in other cultures).

If we conceive of the existential construction in terms of prototypes, then these examples represent the most extreme divergence from a prototypical existential: nothing could be further removed from the requirement of being addressee-new than a pronominal form (except a zero anaphor). In Lambrecht's terms:

> because of the discourse function of presentational clauses, which is to promote brand-new or unused referents to active status, the expressions used to code the "presented" referents are indefinite or definite ACCENTED LEXICAL NOUN PHRASES. Presentational NPs may not normally be pronouns, since the referents of pronouns are already active. (1994: 178)

The only canonical counterpart to this type of construction is the exceedingly rare form *There is that*, which is nevertheless discussed by Birner & Ward (1998: 123–125). There are no relevant examples with this exact phrasing in the whole ICE-GB; the ICE-GB conversation files provide two examples (*In fact there was that* (ICE-GB: S1A-015 #247:1:B); *Right well there was all that* (ICE-GB: S1A-056 #248:4:B), whose discourse meaning, however, is very difficult to ascertain. The following is one of six examples (three spoken, three written) for *There is that* out of the 100 million word *British National Corpus*:

(4.2.25) "Ever thought of a career in the police?" "Blue's not my colour." "I've heard worse reasons." "I can't possibly be tall enough." "They're very flexible about that nowadays." "I've got a degree," I said, getting desperate. "So have I," Prentice checked me. "I couldn't stand the short working week and I really wouldn't know what to do with all that bribe money." "Ah, **there is that**," he said as if thinking it over. "But then, you don't go into the CID straight off ..." I licked a forefinger and made a "nice one" stroke in the air. He was OK but (Rule of Life No. 38) the time to start worrying was when the policemen got nicer. (BNC HTL W_fict_prose)[28]

Birner & Ward treat *there is that* in their section on definite NPs as notional subjects of existential constructions. Recall from above that addressee-new entities introduced via the existential are generally coded as indefinite NPs, so that the definiteness of a notional subject NP requires explanation. Birner & Ward propose that definite NPs in existentials fall into five different classes, but that in "each case, the definiteness of the NP is licensed by the individuability of the referent, while the existential is licensed by its status as hearer-new information" (1998: 122). According to them, the existential sentence *there is that* serves a 'reminder' function (1998: 125): in the example above, the narrator enumerates some arguments why he does not want to become a police officer, finally settling on a sarcastic

28. *CID*: Criminal Investigation Division.

comment. The person called Prentice replies with a thoughtful *Ah,* ***there is that***, as if remembering something from some prior part of the conversation, and it is exactly this "'reminder' flavour" (*ibid.*) which is characteristic of the construction: by referring to the narrator's comment with *there is* ***that***, it is treated "on the one hand as new information but on the other as nonetheless previously known and therefore identifiable" (Huddleston & Pullum 2002: 1398). This analysis clearly does not apply to the frequent IndE phrase *that is there,* where *that* refers to the immediately preceding proposition.[29] The phrase appears as an expression of general consent, indicating agreement with the previous speaker. Pronominal constructions such as *Yes it is there* in example (4.2.23) above similarly articulate support for the previous speaker's proposition, rather than initiating a change of turn by introducing a new referent.

A further large subgroup of examples may be classified as "repetition of the immediately preceding topic NP". There are overall 36 examples in my data, e.g.:

(4.2.26) A: Because now a days **there is** a lot of competition in every field <,,>
B: Competition **is there** but this competition is not healthy <,> often it is manipulated competition <,,> (ICE-IND: S1A-089#141–142)

(4.2.27) F: But when you go to the <.> pr </.> private sector <,> it so happens that <,> they have their own means and methods of testing the students <,,>
See normally <,,> they take for example if it's going to be a salesman's job <,> what they do is <,> they have a group discussion <,> they have an interview
D: No even then <,> **there is** a basic qualification
F: Basic qualification **is always there** <,>
(ICE-IND: S1A-011#168–171)

(4.2.28) A: And do you have social life there <,> in Pilani[30]
B: Pilani
Yes social life **is there** but <,> uh **there are** not so many parties <,> uh as I suppose you have in Goa right (ICE-IND: S1A-065#84–86)

Example (4.2.27) above is interesting in that it features two of the three examples in which the topic NP is introduced by a canonical existential and then taken up by another speaker in a non-initial existential construction. Example (4.2.28) is

29. It might be argued that the demonstrative in the canonical *there is that* in example (4.2.25) above likewise constitutes a link to the immediately preceding utterance; it is, however, difficult to go beyond Birner & Ward's analysis on the basis of so little data of naturally occurring speech. Suffice it to say that the construction is extremely rare in spoken British English.

30. *Pilani*: town in Rajasthan.

similar in that the topic *social life* is repeated in a non-initial existential before the speaker resorts to a canonical existential to introduce a new topic. Most of the examples of the "topic repetiton strategy" (33 examples) simply repeat a preceding NP as the "notional" subject of the non-initial existential, e.g.:

(4.2.29) A: Anyway it's a good experience coming to Bangalore you know
B: Yeah <,>
A: You'll get <,> uh you'll get you get used to <,> spending money
C: No I don't like the crowd know here
A: You know <,,> like you first
C: Something and some places like *Majestic* and all that
A: Yeah *Majestic* **is there**
C: You've a horrid crowd <,> there (ICE-IND: S1A-048#154–161)

(4.2.30) C: Bikaner is famous for its *bhujiya*<,>[31]
A: Bikaner is famous for its *bhujiya* famous for its *rasagullas*
C: *Rasagullas* **is there**
A: Yeah the Bengali *rasagullas* **are there** <,>
(ICE-IND: S1A-008#85–88)

A final noteworthy function that is realized by non-initial existentials with a discourse-old "notional" subject is that of summing up or referring to what has been said before. This function typically occurs without a change of turn between speakers, the speaker him-/herself concludes his/her turn by providing a "summary" of his/her utterance. There are 16 relevant examples in the data, e.g.:

(4.2.31) B: Then uh <,> then uhm how are your colleagues in office <,>
A: Okay okay they are <,>
They'll adjust and they'll help <,> and way of speaking **everything is there** (ICE-IND: S1A-049#181–183)

(4.2.32) A: Apart from the course I'm <,> I like to be in a new place
I like meeting people
B: Yes that's also <,> and <,> that's also **there is** an opportunity we get and we can uh meet <,> so many <,> experienced teachers and we can get the opinion <,> **all those things are there**
(ICE-IND: S1A-029#190–192)

(4.2.33) And there is one latest stock <,> so a lot of conversation <,> jokes <,> anecdotes <,> **they are all there** <,> (ICE-IND: S1A-060#71:1:A)

31. *Bikaner*: city in Rajasthan; *bhujiya*: a spicy snack; *rasagulla*: sweet milk balls (a dessert).

Seven tokens in this group feature a universal quantifier, as in example (4.2.31) above; in example (4.2.32), the universal quantifier precedes the demonstrative pronoun, which in turn modifies the NP. In example (4.2.33), the speaker, a teacher of Russian, first refers to a new textbook for Russian with the canonical existential *there is one latest stock* and subsequently resorts to her list *a lot of conversation <,> jokes <,> anecdotes* with *they are all there,* a construction with a resumptive pronoun as "notional" subject and a postverbal universal quantifier.

To summarize: a substantial proportion, namely 43.3%, of all non-initial existential *there*-constructions in my spoken IndE data feature "notional" subjects that are discourse-old, thus violating the necessary and sufficient condition for a construction to be labelled "existential" in the first place. This should, however, not be taken to mean that the existential is somehow used "wrongly" in IndE: it rather means that we are on the track of an IndE construction whose function does not mirror that of the canonical existential. In order to get closer at a descriptive account of these functions, I will turn to the remaining tokens of non-initial *there*-constructions.

4.2.3.2 *Hearer-new and discourse-new "notional" subjects.* I have described those non-initial existential constructions in my data with a discourse-old "notional" subject in some detail because they represent the most striking divergence from the discourse function associated with the canonical existential, namely to introduce addressee-new referents into the discourse. These examples also raise the question whether the label "non-initial existential" may be a misnomer, since it leads one to believe that the construction would best be analysed with reference to the discourse functions and constraints established for the canonical existential. But before we ponder over a more appropriate label for the construction under scrutiny, we need to examine the remaining examples which, after all, represent the majority of tokens in the ICE-India conversations.

We have already seen that the information status of the "notional" subject does not provide the decisive criterion for the felicitous occurrence of the construction. It should therefore come as no surprise that the remaining 178 examples in the data show both discourse-new and hearer-new "notional" subjects. Since the distinction between discourse-new and hearer-new referents is very difficult to draw indeed in hindsight, even given the contexts of the conversations, I do not want to commit myself to a clearcut distinction between the two categories and provide the respective absolute figures of occurrence. It is safe to say, however, that the hearer-new tokens in the conversation files represent only a minority of around 20 to 30 examples. Example (4.2.6) above truly introduces a new topic. Similarly example (4.2.34): A and B discuss a mutual friend and his plans to go home in

November; the reason for going, *his cousin marriage* ***is there***, represents genuinely new information for A:

(4.2.34) Because his cousin marriage **is there** <,,> that's why he wants to go and be present (ICE-IND: S1A-002#112:1:B)

In the dialogue in (4.2.35) below, A and B discuss a movie, and A provides some details about the movie that are new to B:

(4.2.35) A: I have seen uh *The Hope and Glory* <,> have you seen that <,>
B: No that movie
A: It's a lovely film man <,>
In that you know everything is about nature I'll tell you <,> it's very lovely means very nice lovely <,,> and uh small children **were there** in that (ICE-IND: S1A-053#130–133)

In example (4.2.36), speaker C tells his friends about a bus journey where the driver was driving like a maniac, and he went to the driver to complain:

(4.2.36) In fact I was the first man to go and ask <,> see what you are doing like that
I think some fat people **are there** in the bus <,> he thought it's a big luggage is there in the bus <,> so <,> so instead of thinking that it's a bus he must have thought that it's a lorry (ICE-IND: S1A-017#19–20)

Example (4.2.37) comes very close to being the mirror image of a canonical existential, and there are indeed parallel examples such as (4.2.38) in my data:

(4.2.37) **One thing is there** uh Madhu the <,> this place is blessed with <,,> real good climate <,,> (ICE-IND: S1A-041#159:1:B)

(4.2.38) A: How did you feel <,>
B: Yeah okay but <,> uh but **there's <,> just one thing** which I didn't like <,> like the girls [one word] <,,> (ICE-IND: S1A-059#15–16)

These examples of apparently free variation between the canonical and the non-initial *there*-construction are intriguing as they suggest that there is indeed a case to be made for retaining the label "existential" for the non-initial *there*-construction. In example (4.2.39), speaker B tells A about his hometown Malkhed in Karnataka and its historical significance; in (4.2.40), the conversational setting includes two teachers who are taking part in an English refresher course at CIEFL, Hyderabad. Speaker A compares his course at CIEFL favourably with another refresher course he took in Pune, Maharashtra:

(4.2.39) So <,> even at Malkhed also one *basti* **is there** <,> Jain *basti* <,> at Malkhed <,,>
And uh <,> in that <,> Jain *basti* <,> **there are** some models of uh <,> ancient *bastis* <,> (ICE-IND: S1A-063#88–89)[32]

(4.2.40) In other refresher courses what atmosphere you find <,> atleast you find here <,> after noon sessions **are there** <,> **there are** no afternoon sessions <,>
I did my first at Pune <,> and what atmosphere <,> no refresher <,> so no afternoon sessions nothing like that (ICE-IND: S1A-013#173–174)

The large majority of all remaining examples, however, feature discourse-new "notional" subjects. In example (4.2.41), B tells the story of how her friend wanted to take her out on his motorbike and how, to her great embarrassment, they ran out of petrol. B specifies the location for her listener with *That mosque **is there** no*, and the addition of the particle *no,* seeking confirmation from the listener, indicates that the *mosque* is discourse-new, but represents readily recoverable background knowledge:

(4.2.41) And then we went it stopped [laughter] <,> between and we went pushing you know pushing is really <,> a till there we went <,,> we were pushing till <,,> till Hubli that first petrol pump in Hubli <,> and some <,> I don't no where I think it's just <,> little front of cancer <,> cancer Hospital
That mosque **is there** *no* <,> from there we went pushing and we walked walked (ICE-IND: S1A-040#57–58)[33]

The following excerpt from a conversation between two young students in the English department of Shivaji University in Kolhapur (Maharashtra), provides a host of relevant examples. A and B are planning to organise a fun fair at their university and discuss the details of how to advertise their event across the university campus. All of the subject referents of the non-initial existential constructions can safely be assumed to designate information "which has not been evoked in the current discourse, but which the speaker nonetheless believes is known to the hearer" (Birner & Ward 1998: 15), to repeat the definition from above. The last token of non-initial *there* in the conversational exchange is particularly instructive in that respect: when A notes that their fun fair will take up a lot of their time and that they better stop working for that particular project after December, B readily agrees with *I know studies **are there***, thus voicing a concern that has not been

32. *Basti*: Jain temple; *Jain*: religious community in India.

33. *Hubli*: town in Karnataka, South India; now renamed *Huballi*.

explicitly named before, but which must be uppermost on both speaker's minds and therefore hearer-old:

(4.2.42) A: The thing can be <,> you can have some <,,> what you call the notices <,> and then <,> just go and <,> just put them up in the colleges <,> the different departments of the university physics department **is there** <,> you don't have any interaction with the people from that department <,>
B: Yeah yeah yeah <,>
A: Chemistry department the geography department **is also there**
B: I think we can make it on the university level also like <,> all other departments <,> excluding uh including the humanities and also the science <,> computer this thing we can include all of them <,>
A: But then that will consume lots of time and after December <,> please no more <.> active activities only
B: I know studies **are there**
A: Studies **are there** <,>
This is M A part two with final year <,>
And after this I will be thanking god like anything if I just complete this MA with good marks <,,> (ICE-IND: S1A-062#134–142)

Twelve tokens occur in questions (introduced by either *what* or *how many*). Again these examples do not introduce entirely new referents into the discourse, but take up topics that are already active in the speakers' consciousness. In (4.2.43), sport facilities and upcoming sports competitions have been the topic of the conversation between three young students for quite a while before A complains about the lack of sports facilities in his college. When B asks *Now **what is there** in sports*, he recurs to the current topic of the exchange and just asks for some more specific information:

(4.2.43) A: In our college no sports nothing <,,>
B: Now **what is there** in sports for a
A: Even Badminton **is also not there** <,> in our college <,> only uh volley ball and football (ICE-IND: S1A-051#248–250)

Example (4.2.44) is similar: A and B are two university teachers discussing job-related matters. B is involved in the preparation of a school textbook, and A has previously asked *are you in the committee*? (ICE-IND: S1A-095#227), i.e. the committee responsible for creating and implementing the new textbook.

(4.2.44) How many members **are there** in your committee <,>?
(ICE-IND: S1A-095#246:3:A)

What all these examples again show is that the information status of the "notional" subject cannot be the defining characteristic of the non-initial *there*-construction.

The vast majority of tokens retrieved from the ICE-India conversations represent hearer-old entities, but a small number of all examples actually introduce hearer-new referents into the discourse. It is interesting to note, however, that the non-initial construction ranges over all three possibilities of information status that a referent in discourse may assume. Even if most tokens cluster in the domain of hearer-old information, hearer-new "notional" subjects do occur. These are represented in Table 4.14 in italics. For both the canonical and the non-initial construction, the most frequent function is printed in bold. Moreover, the canonical existential is exclusively associated with the function of introducing hearer-new referents into the discourse: discourse-new referents are only felicitous with presentational constructions (cf. Birner & Ward 1998: 113).

Before finally turning to the canonical existential in the ICE-India conversations, I would briefly like to substantiate the claim noted above that the IndE non-initial construction is associated with a high degree of conceptual orality. A cursory search for the string *is there* in the written part of ICE-India turned up only five relevant examples out of five different text types (two examples from non-printed, non-professional writing and the remaining three from printed sources). Example (4.2.45) comes from a social letter, the second example (4.2.46) from the text type "academic writing/social sciences, and the third example (4.2.47) from a press editorial:

(4.2.45) My College Annual Prize distribution **is there** tomorrow. That will free me of many of my encumbrances. (ICE-IND: W1B-002#89–90)

(4.2.46) Today we have to face reality. Inflation **is there** which means heavy overrun in the [...] cost of various projects necessitating more finances and deficit financing still further. (ICE-IND: W2A-011#15–17)

(4.2.47) Examples of effusive outcomes of banquet bonhomie and dinner diplomacy are ancient, of course. The Sanskrit proverb went: "**Who is there** on the planet who cannot be won over by a full meal? Even a drum makes for mellower music, its faces suffused with salve and powder". (ICE-IND: W2E-007#90–92)

Table 4.14. Comparison of the parameters of information status for canonical and non-initial *there*-constructions

Hearer-new	Hearer-old/discourse-new	Hearer-old/discourse-old
canonical existential	presentational construction	–
non-initial existential	**non-initial existential**	**non-initial existential**

Compared to the figures for the direct conversation files, this type of construction is clearly absolutely marginal in written IndE. However, this simple quantitative statement does not explain anything. Seeking a more principled explanation will be the task of Section 4.2.6 below.

4.2.4 The canonical existential in ICE-India

The canonical existential is twice as frequent as non-initial existential *there* in the ICE-India conversation files, and there seems to be little overall difference in usage compared to other varieties of English. The IndE canonical existential clearly adheres to the discourse condition that only hearer-new referents may occur as notional subjects. Differences occur either at the level of surface realization where other areas of the syntax of spoken IndE are concerned, or as collocational differences. There is, for example, just one single example of an existential tag, as against 16 tokens in ICE-GB:

(4.2.48) A: Do you swim here often <,,>
B: Yeah I go to the well in farm <,,> swimming <,,>
A: No where else <,,>
B: No nowhere else <,>
A: There are also fish in that well <,> **aren't there** <,>
B: Yeah <,> hopefully they were (ICE-IND: S1A-055#140–145)

This is only to be expected, given the marginal role that tag questions play in spoken IndE (cf. Chapter 4.6.2). Another surface difference concerns the presence vs. absence of an indefinite article in the postposed NP. Some of the dozen or so examples are quoted below:

(4.2.49) Sometimes **there is a police** (ICE-IND: S1A-001#48)
(4.2.50) So naturally **there will be clash** and uhm <,> our democracy <,> would wouldn't be able to survive (ICE-IND: S1A-005#28:1:B)
(4.2.51) **There is lot of scope** for <,> uh research <,> (CE-IND: S1A-063#55)

The expressions *there is/are (a) lot of/sort of/number of X* are particularly prone to variation with respect to the feature "+/- indefinite article", a feature that has been extensively studied by Sharma (2003, 2005), among others. However, the absence of indefinite articles in examples (4.2.50) and (4.2.51) above does not impinge on the definiteness of the NPs, which are fully felicitous as hearer-new referents.

An intriguing difference concerns the number of bi-clausal existentials, which is lower in ICE-India in comparison to ICE-GB (the difference is statistically significant according to the log-likelihood test), e.g.:

(4.2.52) No but marriage is just a part of <,> uh life you know but when we're getting ourselves educated **there is no point sitting at home** as such doing Ph D's and all that <,> hey you can just sit with your BA or
(ICE-IND: S1A-011#56:1:C)

This and other aspects such as *no*- vs. *not*-negation in canonical existentials in the two varieties remain to be pursued further. However, since they are of no immediate relevance here, I will leave them for future studies.

4.2.5 Initial and non-initial existential constructions in spoken IndE: Speaker variables

Before I turn to an analysis of the data in correlation to speaker variables, it is worth summarizing the findings on the discourse properties of the two existentials. The canonical existential in spoken IndE is no different from other varieties of English in that its most pronounced function is to introduce a new referent into the discourse, which is then made available for further predication. The NP designating the referent appears in the postverbal focus position, which is the preferred position for new information, and the subject/topic position is taken up by nonreferential *there*. For this discourse function, the canonical existential is unambiguously the majority option in spoken IndE: only a small number of tokens of the non-initial existential *there*-constructions likewise introduce new referents. The two constructions overlap only minimally with respect to this particular discourse function because, strictly speaking, they are not variables in the sense of variable realizations of one and the same function.

The main function of the non-initial construction, in turn, seems to be the assertion of existence or occurrence of a referent that may readily be retrieved from the discourse context, i.e. a referent that is hearer-old or even discourse-old. To assert the existence or occurrence of an already established referent may seem redundant, even paradoxical, but it is exactly this redundancy which motivates the construction in spoken IndE: we will see again in the following sections that repetition is an important strategy for creating cohesion in IndE discourse. I will return to this point below in trying to account for the structure of the non-initial existential *there*-construction.

The data for the speaker variables in correlation with the occurrence of the two existential constructions afford some interesting observations. Gender is clearly a significant factor with respect to the likelihood of occurrence of either the canonical or the non-initial construction (at $p = < 0.0001$). The distribution of the two constructions with female speakers is almost a mirror image of the pattern with male speakers, as Figure 4.15 illustrates. Men's use of both constructions is

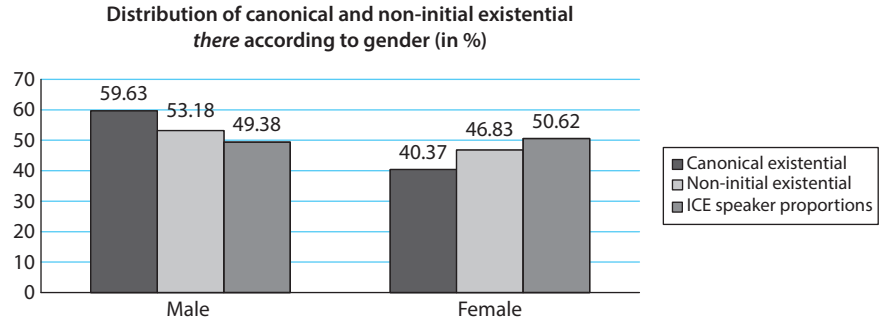

Figure 4.15. Distribution of canonical and non-initial existential *there* in the ICE-India conversation files according to gender

proportionately higher than that of women, with an even greater gap between the sexes in the occurrence of the canonical existential.

The figures in relation to speakers' age are equally significant, but do not present a general tendency comparable to the data for speakers' gender. The youngest age group (18–25) clearly favours the non-initial construction, and the oldest age group (50+) equally clearly shows a preference for the canonical existential, but the intervening age groups present a mixed picture. The non-initial construction has another peak in the age group of the 34–41 year olds.

The data for the variable "educational level" do not lead to a clearer picture, either, even though they show again a statistically significant distribution (at $p = 0.0007$): the non-initial existential is under-represented with speakers who hold a PhD, but it is similarly underrepresented with students in secondary education.

Finally, the distribution of the two existentials with respect to language family is not statistically significant (at $p = 0.4941$).

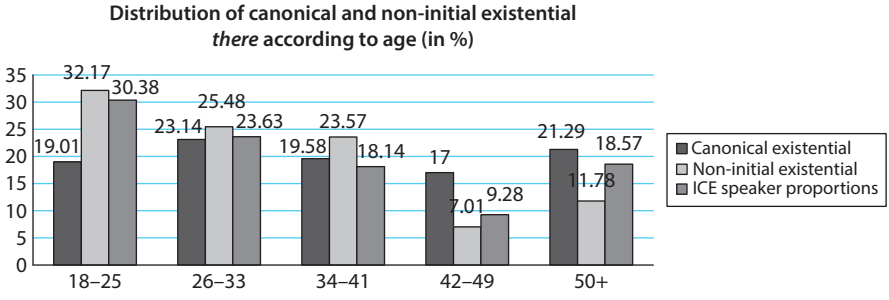

Figure 4.16. Distribution of canonical and non-initial existential *there* in the ICE-India conversation files according to age

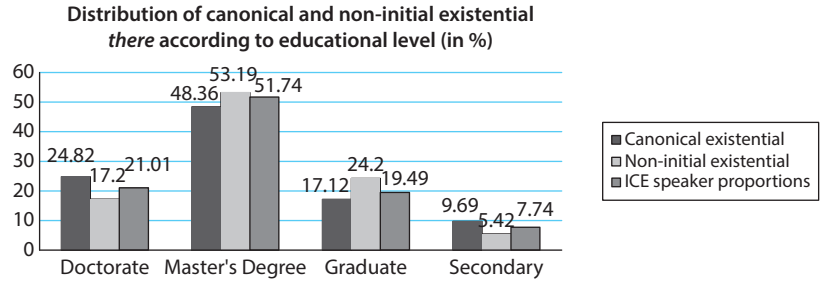

Figure 4.17. Distribution of canonical and non-initial existential *there* in the ICE-India conversation files according to educational level

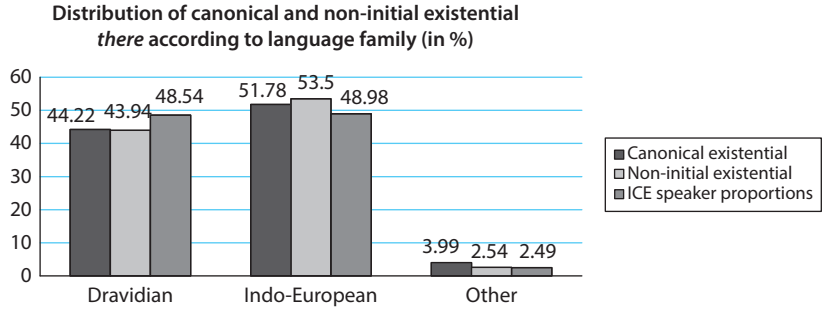

Figure 4.18. Distribution of canonical and non-initial existential *there* in the ICE-India conversation files according to language family

Thus, the independent variables age, gender, education etc. do have a bearing on the choice of one construction over the other, but it is only with respect to the variable gender that there is also an inverse relationship between the preferences for one construction over the other. For the variables age and education, we find tendencies within the individual groups, but these tendencies do not add up to an overall trend or pattern. A much more detailed analysis, based on figures for the different discourse functions of the existential constructions, would probably be required here.

However, we might venture the conclusion that the non-initial existential construction is fairly well entrenched in the IndE speech community and that it is no less "standard" in spoken IndE than the canonical existential. Moreover, a plain quantitative sociolinguistic analysis along the lines of age would potentially lure us to misleading conclusions: if we only looked at the peak for the non-initial construction in the youngest age group and the corresponding peak for the canonical existential in the oldest age group, we might have been inclined to evoke the standard sociolinguistic explanations which correlate "young" with "innovative,

informal, nonstandard" and "old" with "conservative, correct, standard". However, the very fact that the non-initial construction is underrepresented both with the "doctorate" group *and* the "secondary (undergraduate/college)" group forecloses any such interpretation.

Still, the virtual absence of the non-initial construction from written IndE has to be explained, and the explanation is more likely to be related to the parameters of orality vs. literacy than to speaker-related variables such as education etc. Returning to the investigation of the discourse motivation of the non-initial *there*-construction will also take us closer to an explanation for the origin of the construction in multilingual discourse.

4.2.6 Explanatory parameters

It was the apparent structural parallelism between canonical existential *there* and non-initial *there* which prompted my data-based comparison in the first place. The preceding investigation has revealed that there is no corresponding functional parallelism between the two constructions. In the end, this does not come as a surprise, as I have already indicated above, yet it is perhaps more surprising that the non-initial construction does indeed occasionally introduce a new discourse referent rather than being restricted to taking up hearer-old referents. In its dominant function, the non-initial *there*-construction is then more akin to topicalization in spoken IndE, as the next section will show.

Having accounted for the function of the non-initial *there*-construction, we might still wonder about its provenance – after all, the construction is exclusively found in IndE. "Substrate influence" is generally the first choice when having to explain a linguistic feature which is unique to a particular New English variety. I will therefore turn to existentials and related constructions in Indian languages in order to investigate to what extent the notion of 'substrate influence' can be enlisted as an explanatory parameter.

4.2.6.1 *Existentials in Indian languages.* As topic-prominent languages (cf. Chapter 3.1.3), Indian languages generally have no dummy subjects equivalent to English *it* or *there*. Word order in Indian languages is also more flexible than in English: postposing of constituents which designate new discourse referents can be achieved just by a change in word order. In McGregor's terms: "the sense of introductory 'there is, are, were' etc. is expressed in Hindi by an inversion of the neutral (non-affective) order of subject and following adverbial locution (emphasis on the subject being thereby diminished)" (1995: 4). His examples are (gloss added):

(4.2.53) *pustak mez par hai* (book-table-on-is)
The book is on the table.
mez par pustak hai (table-on-book-is)
There is a book on the table. (*ibid.*)

Examples such as these are termed "existential expressions with Topicalized Locatives" by Masica (1993: 359) with reference to the Indo-Aryan languages. In Hindi as in the other Indo-Aryan languages, the focus of the sentence generally precedes the finite verb. Kachru provides the following example from Hindi to illustrate how a "new entity is introduced in the focus position and subsequently becomes the theme" in narratives:

(4.2.54) kisī zəmane mẽ ek **raja** t^ha.
some time.M.OBL. in a king be.PAST.M.SG.
uske do raniyã t^hī.
he.POSS.OBL. two queen.F.PL. be.PAST.F.PL.
'Once upon a time there was **a king. He** had two queens.'
(Kachru 2006: 252)

The Dravidian languages are no different in that respect, as the following example from Malayalam illustrates:

(4.2.55) meeʃa meel pustakam unʈə
table on book be-PRES
"There is a book on the table." (Asher & Kumari 1997: 100)

Asher & Kumari further note that for "the existential 'being' verb *unʈə*, except when the location is universal, as in *daivam unʈə* 'God exists', it is necessary to state the location, unless this has already emerged from the context" (1997: 115). Again, this is similar to Hindi and the other Indo-Aryan languages, where an assertion of existence is a speech act with a rather restricted range:

(4.2.56) īʃvər hɛ.
God be.PRES.SG.
'God exists.' (Kachru 2006: 190)

What English and the Indian languages have in common, then, is the syntactic strategy of postposition to assign salient discourse referents to the position for new information in the sentence. In a strict SVO language such as English, the postposed constituent appears after the verb, but preceding any other complements or adjuncts, and the subject position has to be occupied by a nonreferential element. In Indian languages, which are almost all SOV, the postposed constituent occurs before the finite verb, and complements or adjuncts appear in

sentence-initial topic position. The relevant structure across Indian languages is then: 'topic (usually locative NP) – focus NP/postposed constituent – form of *be*'. The Hindi equivalent of BrE *There's food (in the fridge)* would then be *(fridge me) khana hai* ('(fridge – in) – food – is'), and the IndE equivalent *food is there*.[34] Again, the structural pattern 'focus NP + be' would be the same in (almost) all Indian languages. What happens, then, is that speakers make use of a pattern that is very common in their mother tongues and transfer it to English. A literal translation, or calque, is ruled out for two reasons: in both English and Indian languages, a sentence such as *food is/khana hai* would be used to assert the existence of some entity in a philosophical sense, a rather awkward speech act when applied to *food*. In English, "*be* can hardly occur without an internal complement" (Huddleston & Pullum 2002: 1393), and in order to avoid a violation of English surface syntax, IndE speakers complete the gap in the complementation pattern of *be* with dummy *there*. Here, the constraint on pattern replication observed by Matras & Sakel becomes apparent (2007: 830f.):

> Note that the procedure which we call pattern replication operates under the constraint of the exclusion or avoidance of direct replication of matter from the model language. This means that overtly, the structural coherence of the replica language as the chosen language of the communicative interaction (at least in respect of the utterance or the construction in question) is respected. The procedure is to replicate the abstract organisational pattern of the model construction using suitable elements in the replica language. Strictly speaking, we are therefore dealing with the export of constructions from a model language to the replica language, rather than with an import or 'borrowing' of structures: the speakers' motivation is to avail themselves of constructions that are part of their overall linguistic repertoire, irrespective of the setting of the interaction (and so irrespective of the choice of a particular language 'system' for the current interaction). In order to do this, and at the same time to respect the overt structural coherence of the chosen language of the ongoing interaction, speakers turn to the creative process of pivot-matching.

Matras & Sakel's notion of pivot-matching has been treated in more detail in Section 2.2.3 above. The "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.

Conceiving of the non-initial *there*-construction as an "export" rather than as an "import" also helps to explain the difference in information status between the non-initial and the canonical construction: the Indian construction 'focus NP + *be*' requires the focus NP to designate salient information in the ongoing

34. I thank Kasturi Dadhe for this example and for her extremely helpful comments on data from Hindi, Marathi, and of course Indian English.

communicative situation. Such information is not necessarily hearer-new, especially not in discourse, where repetition is very common as a cohesive device:

> The process of repeating to stress emotional agreement seems highly conventionalized in Indian languages. [...] Elliptical repetitions help to please the positive face of the hearer and to reduce any uncertainty. This strategy is so common that often the interlocutor's repetitions may not be relevant to the point being made, hence for a non-Indian hearer, the Indian speaker is viewed as repetitive and inconsistent. (Valentine 1995: 233)

Thus, the IndE non-initial existential construction observes the discourse constraints on information status as derived from its model or source, rather than from its target. The intriguing question then remains why *there* was appropriated for this particular construction. Matras & Sakel's theory as outlined in Chapter 2.3 is again instructive on this point:

> the grammaticalisation process begins by matching lexemes to one another and adapting the range of meanings expressed by the lexemes of the replica language to those expressed by the parallel lexemes in the model. The basis for the matching procedure is the polysemy of the word in the model [...] which represents both a more concrete and a more abstract meaning, [...] Speakers will inevitably direct their attention to the more concrete meaning when searching for a match in the replica [...]. The match will then lead to the emergence of a more abstract meaning. (Matras 2009: 238f.)

A speaker of IndE might be guided by the canonical existential construction to select nonreferential *there* as the missing complement for the non-initial construction, but I would maintain that the non-initial construction is not simply a reversed canonical existential: this hypothesis would obscure the actual process of pivot-matching that is at work here. It is the polyfunctionality of *there*, and not its occurrence in a highly specific context such as the canonical existential, which is exploited creatively by IndE speakers. As noted above, the polysemy of *there* includes the adverbial (locative/deictic) as well as the more abstract nonreferential meaning, and the more basic locative meaning is still apparent in bare canonical existentials, as Huddleston & Pullum have noted: their example *there is plenty of ice-cream* "illustrates the common case where *there* is an implicit locative" (2002: 1393). An utterance such as *there is plenty of ice-cream* triggers the implicature that the ice-cream is somewhere close by.

The close connection between the notions of existence and of location is also apparent in the grammaticalization of *there* from deictic adverb to nonreferential marker of the existential construction, a grammaticalization trajectory which is found in many other languages as well (cf. French *il y a*; Lambrecht 1994: 180). The

following example illustrates the polysemy of *there* and the subtle transition from locative to non-referential final *there*:

(4.2.57) C: Oh well are you happy here <,,> Satyam?
B: It's too early to say that but so far it's fine <,>
And <,> uh I've to be careful of the mosquitoes here [laughter]
And uh I think it should be fine and <,> I had long desired to come here because <,,> uh it is the what you say Mecca for <,,> uh English
C: English
A: You you don't have mosquitoes **there** in your home <,>?
B: No <,,> we have mosquitoes **there**
But <,> mosquitoes of CIEFL are famous [laughter]
C: I mean that what you told <,,>
B: Uh you see the mosquitoes **are there** <,> mosquito curtains **are not there** and uh <,,> (ICE-IND: S1A-013#46–50)

The two teachers who meet at CIEFL in Hyderabad for an English refresher course have already featured in example (4.2.40) above, in this excerpt, they are discussing the abundance of mosquitoes on the CIEFL campus. When speaker A inquires after the presence of mosquitoes with the utterance *you don't have mosquitoes **there** in your home?*, *there* is clearly locative. B's answer *No we have mosquitoes **there*** first of all illustrates how "'no' means 'yes'" (Valentine 1995: 243) and continues to use *there* in a locative meaning, echoing A's question. The transition from locative to nonreferential occurs in B's next turn with two fully-fledged non-initial constructions: *the mosquitoes **are there** <,> mosquito curtains **are not there***. Speaker B here quite literally proceeds from the "more concrete, semantically 'basic'" (Matras & Sakel 2007: 852) meaning of *there* to the more abstract, grammaticalized meaning.

The non-initial existential *there*-construction in spoken IndE, then, is neither a mirror image of the canonical existential nor an alternative to it, but an addition to the repertoire of spoken IndE with largely different discourse functions. In its emergence in mutilingual discourse, specific linguistic properties related to India as a discourse area conspire with properties related to India as a sociolinguistic area: a structural pattern common in the Indian SOV languages is re-created in English while preserving the structural integrity of English. In discourse, the non-initial construction carries much of the burden of repetition and establishing cohesion, discourse strategies that are closely associated with the South Asian sociolinguistic area (cf. Section 3.1.4 above).

To the best of my knowledge, the non-initial construction is not stigmatized or somehow considered "bad English" by members of the IndE speech community (unlike the pragmatic marker *no/na* and the focus marker *only*, cf. the relevant

sections below), and the data of the distribution of the form in the IndE speech community discussed above support this. Non-initial *there*-constructions are fully compatible with English syntax: they do not violate any grammatical rule in English, and unlike *only*, they are not transparently borrowed or calqued, but created by selecting an appropriate pivot in the replica language. Further, the construction may not even be obvious as an innovation because it patterns with the presentational construction in which *there* is deictic/locative, as in the example discussed above. If the construction does not cross the threshold from spoken standard IndE to written standard IndE, then this is not because of some kind of stigma attached to it, but because its function is firmly associated with conceptual orality and the "language of immediacy" (Koch & Oesterreicher 2012). Written language simply has no need for the construction.

4.3 Topicalization

New Englishes generally are credited with a marked preference for topicalization constructions, "in which a non-subject is 'topicalized', i.e. marked as a topic expression by being placed in the sentence-initial position normally occupied by the topical subject" (Lambrecht 1994: 147). Despite its putative prominence, topicalization is not represented in Kortmann & Szmrecsanyi's (2004) global synopsis of recurrent features across varieties of English, and empirical studies on the subject are few: exceptions are Mesthrie (1992: 110–127) on South African Indian English and Mesthrie (1997) on Black South African English. For IndE, the subject has been mentioned in passing. Sharma (2003: 56) found that "discourse-driven non-canonical word order" occurred frequently in her data, a feature she attributes to "direct L1 transfer" (2003: 59). Bhatt (2008: 553), in turn, notes that topicalization is fairly common in IndE. As such, spoken IndE is no exception to the general tendency of vernacular Englishes to make higher use of discourse-pragmatic word order. Bhatt does not find any restrictions on the kind of constituent that may be topicalized. Object noun phrases are the prime candidates for topicalization, but adverbials of place and time are also frequently fronted (2008: 553). "However, the more surprising aspect of the syntax of topicalization in vernacular IndE is that it is fairly widespread even in embedded contexts" (2008: 554).

To summarize the state of the art so far:

> Topicalisation phenomena are common in informal Std Eng [Standard English]. In New Englishes they are noteworthy for several reasons: (i) the extension to a wider range of grammatical contexts; (ii) a (possibly) higher frequency of occurrence; and (iii) an extension to formal contexts. (Mesthrie & Bhatt 2008: 81)

Before I proceed with the analysis of relevant data from ICE-India, it is necessary to arrive at a more precise definition of what constitutes topicalization in the present context and thus to delimit the range of phenomena that will be dealt with in this chapter. I will then discuss in some detail the topicalization constructions in ICE-GB as a basis for comparison with topicalization constructions in ICE-India – not least because the ICE-GB data diverge in interesting ways from some descriptive accounts that have been offered so far.

4.3.1 Definition

When Mesthrie & Bhatt name topicalization as a noteworthy feature of New Englishes (cf. also Mesthrie 2008: 633), they take topicalization as a cover term for left dislocation, cleft constructions, and *fronting*, "in which an NP is fronted, without an overt trace in the main clause" (2008: 82). Topicalization is indeed "superficially similar" (Lambrecht 2001b: 1052) to left dislocation, as both involve

> a referential constituent in non-canonical initial position. But while in LD [Left Dislocation] this constituent occurs in the extra-clausal TOP position, the "topicalized" phrase occurs in the so-called pre-clausal COMP (complementizer) or WH-position, where it preserves its syntactic and semantic role as a complement of the verb. (*ibid.*)

I will reserve the term "topicalization" for what Mesthrie & Bhatt term "fronting" and treat the related phenomena of left dislocation and clefting separately. Other scholars refer to the construction as *preposing* (e.g. Huddleston & Pullum 2002: 1372, Ward & Birner 2004: 158–160):[35]

> a Preposing is a sentence in which a lexically governed, or subcategorized, phrasal constituent appears to the left of its canonical position, typically sentence-initial. Preposing is not restricted to any particular phrasal category [...]. In each case, a single argument appears in preposed position and [...] that argument is constrained to be old information. More specifically felicitous preposing in English requires that the information conveyed by the preposed constituent constitute a discourse-old anaphoric link to the preceding discourse. (Ward & Birner 2004: 158f.)

Note that Ward & Birner's definition differs from both Mesthrie & Bhatt's and Lambrecht's characterization in two respects. First, the preposed constituent need not be an NP, e.g.:

35. Ward & Birner (2004: 160) distinguish two types of preposing, namely focus preposing and topicalization. Huddleston & Pullum (2002: 1376–1381) similarly make a distinction between non-focus complement preposing and focus complement preposing. I will return to these distinctions below.

(4.3.1) Oh **great** that'd be (ICE-GB: S1A#042: 132)
(4.3.2) **Distinctly brilliant** I thought him (ICE-GB: S1A-094#312:1:A)

In both examples, adjective phrases are topicalized: in (4.3.1) the adjective is the subject complement, in (4.3.2) the object complement.[36] Secondly, the definition requires the preposed constituent to be an argument of the verb, thus excluding adverbials as e.g. in the following examples:

(4.3.3) **just like any other dance group** we would be self-financing <,> (ICE-GB: S1A-001#097:1:B)
(4.3.4) **In England** it doesn't matter because <,> he can get help as he's uh the only one allowed to drive that car <,> (ICE-GB: S1A-009#232:1:B)
(4.3.5) **In the summer** you can take a car and four people for a hundred and twenty pounds (ICE-GB: S1A-021#105:1:C)

Huddleston & Pullum clarify the pragmatic motivation of this constraint:

> Adjuncts may also be located in prenuclear position, but when complements occur there the construction is subject to pragmatic constraints on familiarity status that do not apply with initial adjuncts. (2002: 1372)

An example like (4.3.5) ***In the summer** you can take a car*... does not need any special context to be felicitous, whereas their contrasting example **In a basket I put your clothes* "requires a context that motivates the preposing" (*ibid.*). Moreover, the constraint that only arguments can be topics and/or moved to topic position is a logical consequence of the very definition of 'topic': since the topic instantiates what the proposition is 'about', it is intuitively obvious that adjuncts cannot be topics. I will therefore follow Ward & Birner in considering only those cases where an argument of the verb is placed in sentence-initial position. In this, I diverge from Mesthrie (1992, 1997), who included fronted locative and temporal NPs as instances of topicalization when they appeared to be marked. As Mesthrie elaborates, markedness judgements in this realm are inevitably made on an intuitive basis (1992: 116–117). He does not include an example like (4.3.6) because it represents "a regular feature of most English dialects", whereas he considers (4.3.7) to display a "high degree of markedness":

(4.3.6) Every Monday we go for service.
(4.3.7) You see, morning I was reading in the *Post* too. (= 'I read it in the *Post* this morning') (*ibid.*)

36. Cf. Kiss (2001: 1444f.) for a discussion whether preposed predicative adjectives should be analysed as contrastive topics or rather as a focus construction. The argument that predicative adjectives cannot be topics concerns mainly theory-internal issues pertaining to generative grammar and need not be spelled out here.

In order to avoid making unwarranted intuitive judgments, I will restrict the following discussion to topicalization in a narrow sense, i.e. the fronting of arguments. However, some interesting related cases will be taken up in Section 4.3.4 below.

Researchers unanimously agree that the topicalized constituent has to convey information that has already been evoked in the discourse context. "[F]elicitous preposing in English requires that the information conveyed by the preposed constituent constitute a discourse-old anaphoric link to the preceding discourse" (Ward & Birner 2004: 159), as in the following example where the topicalized constituent is a discourse-deictic *that*:

(4.3.8) Uhm <,> I mean I'm applying this week for a job at the V and A and I've already been told there's a strong internal candidate <,,> in the India department and **that** I was told by the head of the department
(ICE-GB: S1A-066#057–59:B)

The following lengthier stretch of discourse from a speaker who talks lovingly about his book-hunting in secondhand shops may serve to illustrate another frequent characteristic of topicalization constructions:

(4.3.9) Oh well a lot of these books have been out of print for <,> well not
Yeah
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#082–87:E)

In this context, *hardbacks* are contrasted with *paperbacks*, and even if *hardbacks* had not been explicitly mentioned by the speaker, the listener would have had no problems in recovering the link between the overall topic of the conversation, namely *books*, and a subset of these, the *hardbacks*. It has become customary to describe such possible links as constituting partially ordered sets, or "posets":

> *Discourse Condition on Preposing*
> The entity represented by the preposed constituent must be related, via a salient partially ordered set relation, to one or more entities already evoked in the discourse. (Ward & Prince 1991: 173)

Poset relations pertaining to arguments may include "such relations as type/subtype, entity/attribute, part/whole, identity, etc." (Ward & Birner 2004: 159). This

discourse condition applies to both types of preposing, namely topicalization as in the example above as well as focus preposing, as in (4.3.10):

(4.3.10) A: I picked up a load of uniforms
They're not very nice
B: Oh yeah yeah
Do they fit and uhm
A: Well they're all right but they've got uhm <,,> two pockets on the front not one
B: What
on the top <,>
High pockets you mean
A: Yeah (ICE-GB: S1A-099#142–150)

The sentence ***High pockets*** *you mean* has just one nuclear accent on the fronted focus constituent *high pockets* and it is related to the preceding discourse by virtue of being in a type/subtype poset relation with the aforementioned *pockets*. The sentence *But uh* ***hardbacks*** *I wouldn't lend to anyone,* on the other hand, contains two accents: one on the fronted topic *hardbacks* and another on the focus *anyone*. In ***that*** *I was told by the head of the department* in example (4.3.8) above, anaphoric ***that*** establishes an "identity link to the prior discourse" (Ward & Birner 2004: 160), i.e. the relevant poset consists of only one member. The requirement that the preposed constituent must form "a discourse-old link to the preceding discourse via a salient linking relation" (2004: 162) holds for both types of preposing. In the following, I will only differentiate between topicalization (in Ward & Birner's narrow sense) and focus preposing when it impinges upon my argument. There is, however, one apparent paradox that needs to be resolved: how can the focus of a sentence, which by definition contains what is addressee-new or unpredictable information, satisfy the requirement of constituting a discourse-old link to the preceding discourse when preposed? Huddleston & Pullum elaborate on this problem and offer the following example for discussion:

(4.3.11) A: Did they give the job to you or to Mary?
B: They gave it to HER. (2002: 1370)

In B's answer to A's question, the pronoun refers to Mary and therefore constitutes discourse-old information; on the other hand, the fact that it was Mary who got the job is new to the hearer. As Huddleston & Pullum argue:

> This apparent paradox can be resolved by distinguishing between the familiarity status of Mary considered as an entity, a person (from this point of view Mary is old information), and Mary considered as the value assigned to the variable in the focus frame "they gave it to x" (from this point of view Mary is new information). (*ibid.*)

Another example from ICE-GB which actually involves focus preposing may serve to clarify the distinction. The conversational setting includes a group of people sitting around a dinner table and discussing dessert options. The choice is between different flavours of ice-cream, meringue, or both. C's comment is about a person who is at the moment absent from the conversation:

(4.3.12) **Just ice cream** he wanted (ICE-GB: S1A-073#316:1:C)

Here again, the preposed focus *Just ice cream* is clearly discourse-old, as the discussion has been turning around vanilla ice cream and strawberry ice cream for some time. As the value which instantiates or specifies the variable y in the open proposition "person X wants y for dessert", *Just ice cream* conveys new information by excluding other possible options in the given context. In Ward & Birner's terms (2004: 161):

> While the anchoring poset [...] is discourse-old, the preposed constituent itself represents information that has not been explicitly evoked in the prior discourse. In the case of focus preposing, then, the related poset must be discourse-old while the link – as focus – is new. Thus, it follows that the poset must contain at least one other member in addition to the link.

This means that identity links or "singleton posets" such as discussed above may occur in topicalization proper, but not in focus preposings. We will come back to this point below.

The discussion of focus preposing raises another question which has been sidestepped in most of the available literature on topicalization: much has been written on discourse constraints, but to the best of my knowledge, no-one except Lambrecht (1994) has looked beyond discourse constraints for discourse *motivations* of topicalizations. The question of motivation is relevant for both topicalization proper and focus preposing, but becomes particularly prevalent in the context of focus preposing: since the focus of an utterance is by definition already made prominent by a pitch accent, the question arises why the focus constituent should receive additional syntactic marking by being placed in preverbal position. As Lambrecht – to me, correctly – states: "Cognitive accessibility is only a necessary, not a sufficient condition for the use of a construction such as topicalization" (1994: 163).

The following excerpt from a conversation about making pastry dough may serve to illustrate the problem that arises with definitions which limit themselves to stating that the topicalized constituent must provide a discourse-old link:

(4.3.13) B: Well with the Barry technique with the processor you process the dry things together and the butter <,> and then you put the water and the eggs in with a quick wiz at the end
A: Mm
Mm <,>
But <,> uhm **one thing** I found
and people have told me this which is that if you <,,> put a very small amount of water in and if you then <,> wizz it for longer the same amount of water a small amount of water does the binding for you whereas if you only wizz it for a short time you know then it ends up looking crumbly and you think
ah more water (ICE-GB: S1A-057#180–185)

B, a male retired engineer, age 66+, is holding forth to A (a female lecturer, age 26–45) about the correct way of making pastry dough. Speaker A first corresponds with a series of *mms* before she takes the floor from A and describes an alternative to the aforementioned "Barry technique". A's utterance ***one thing** I found* provides a rather weak link to the preceding discourse; ***one thing*** may be understood as belonging to a poset of generic 'things relevant to pastry making'. As I see it, the most important function of A's focus preposing is topic-shift: A tries to take the turn from B. As Lambrecht put it (1994: 161):

> the relevant function of topicalization is not to mark an activation state of a referent but to mark the referent of an NP as a (particular kind of) TOPIC in the proposition in which it is an argument and, as a corollary, to mark the proposition as being about the referent of that topic. Such syntactic marking is necessary because in sentences with unmarked presuppositional structure accented object NPs are not topics but focus constituents.

A might just as well have said *There's **one thing** I found*, or *I found* ONE THING, with a pitch accent on the focus *thing* in its canonical object position. In ***one thing** I found,* the NP is a contrastively focussed topic introducing an attempt at topic-shift. Lambrecht's characterization of the discourse motivation of topicalization thus provides a more insightful account than most other scholars' insistence on discourse constraints.

The preceding descriptive generalizations concerning topicalization in English are derived from research on British and American English. For the time being, I will assume that topicalization in IndE follows the same or similar discourse constraints, and focus on the form and frequency of this construction in order to examine Mesthrie & Bhatt's claim quoted above. However, Mesthrie (1992) has provided clear evidence for a wider range of discourse functions of topicalization constructions in SAIE, prompting us to look more closely at our IndE data in a subsequent section.

4.3.2 Topicalization in ICE-India and ICE-GB: Form and frequency

The topicalization constructions in ICE-GB were retrieved by searching for e.g. the inexact node ***preod*** (preposed direct object) and then excluding those examples where e.g. the preposed direct object was realized by an interrogative.[37] Table 4.19 provides the token frequency of topicalization constructions in ICE-GB and ICE-India.

The difference in frequency between the two varieties is striking: we find almost ten times as many topicalization constructions in IndE as compared to BrE. Gregory & Michaelis (2001) provide comparable data from the Switchboard corpus of AmE telephone conversations with a frequency of 17.6 instances of topicalization per 100,000 words. Netz & Kuzar (2007) found only six tokens of topicalization in the first part of the Santa Barbara Corpus of Spoken American English.[38] Mesthrie & Bhatt's claim that topicalization phenomena show "a (possibly) higher

Table 4.19. Topicalization tokens in the conversation files of ICE-India and ICE-GB (percentages as of the overall number of tokens)

Topicalization	ICE-India (tokens)	ICE-GB (tokens)
direct objects	**147 (78.6%)**	**12 (63.15%)**
noun phrases	114 (60.96%)	7 (36.84%)
clauses	4 (2.14%)	1 (5.26%)
anaphoric pronouns	29 (15.5%)	3 (15.79%)
other	–	1 (5.26%)
subject complements	**38 (20.32%)**	**5 (26.31%)**
noun phrases	17 (9.1%)	2 (10.52%)
adjective phrases	20 (10.7%)	2 (10.52%)
anaphoric pronouns	1 (0.53%)*	–
other	–	1 (5.26%)
object complements	**2 (1.07%)**	**2 (10.52%)**
Total	**187**	**19**
per 100,000 words	**85,57**	**9,24**

* The example is: *like that he is....*

37. The relevant inexact nodes were *preod* (preposed direct object), *preoi* (preposed indirect object), *precs* (preposed subject complement) and *preco* (preposed object complement). For ICE-India, the topicalization constructions were retrieved by simply reading all files and coding relevant tokens manually.

38. The authors do not provide information about the size of their subcorpus, nor does the homepage of the *Santa Barbara Corpus of Spoken American English*. The whole corpus (Parts 1–4) comprises about 249,000 words; each part may then consist of 40,000 to 50,000 words, so that we may guess at a frequency of 12–15 tokens per 100,000 words.

frequency of occurrence" (2008: 81) in New Englishes is thus amply confirmed for spoken IndE. Since our data come from standard spoken IndE, we may surmise that the third part of their statement which refers to the constructions' "extension to formal contexts" (2008: 81) is also borne out.[39] Their first point about topicalization in New Englishes, namely the occurrence in a wider range of grammatical contexts, is exemplified by Bhatt (2008) for IndE. As already mentioned above, he found topicalization constructions even in embedded contexts, e.g.:

(4.3.14) His friends know that **her parents**, he doesn't like at all.
Papa-ji only told us that **their money**, he will not touch
My brother warned me that **young boys**, I should say no to.
(Bhatt 2008: 554; emphasis mine)

The occurrence of embedded topicalization constructions in ICE-India is of course not apparent from our table. However, a look at the 187 examples retrieved from ICE-India does not provide conclusive support for the extension of the construction to the kind of contexts that Bhatt is envisaging.[40] Examples (4.3.15) to (4.3.19) constitute an exhaustive list of all topicalization tokens in complex sentences:

(4.3.15) I have been to a few places and <,> uh and **the remaining few** I plan to see later <,> (ICE-IND: S1A-014#19:1:B)

(4.3.16) The department is now thinking <,> **the operators** to absorb in some other departments <,> and uh they'll be posted to some to some booths <,> or <,> uhm they'll be taken as clerks <,> in the in the various departments <,> that is going on <,,> (ICE-IND: S1A-018#171:1:B)

(4.3.17) C: If she has a passport it's well and good otherwise great problem *yaar*
A: Yeah
That she has I think (ICE-IND: S1A-037#245–247)

(4.3.18) A: She is getting nightmares <,> dreadful dreams
B: Uhm
Haan haan
A: **This** she is getting it seems <,,>
And she wanted to contact the doctor about it <,,>
(ICE-IND: S1A-052#71–75)

(4.3.19) Then he said **another test** we will take (ICE-IND: S1A-068#147:1:A)

39. To be sure, a full analysis of "formal contexts" would have to include topicalization constructions in written IndE and written BrE.

40. It is not apparent from the context of Bhatt's article whether the examples he provides for embedded topicalization come from his corpus data of actual IndE conversations or whether they are derived from introspection.

In (4.3.15) and (4.3.16), the topicalized constituents are moved out of a *to*-infinitive clause. The non-topicalized counterparts to examples (4.3.17), (4.3.18) and (4.3.19) could represent embedded contexts along the lines of Bhatt (2008: 554):

(4.3.17) a. I think (that) she has **that**
(4.3.18) a. It seems (that) she is getting **this**
(4.3.19) a. Then he said (that) we will take **another test**

We might then expect that topicalization produces surface structures similar to those quoted from Bhatt above:

(4.3.17) b. I think (that) **that** she has
(4.3.18) b. It seems (that) **this** she is getting

However, this is not what we get: in all examples, topicalization moves the relevant constituent to clause-initial position. Example (4.3.16) differs from the others in that the topicalized constituent introduces the *to*-infinitive clause which is syntactically dependent on the preceding matrix clause. In examples (4.3.17) and (4.3.18), the topicalized constituent introduces the subordinate clause which precedes the matrix clause. Example (4.3.19) comes closest to Bhatt's examples; however, since the subordinator *that* is omitted, we have no way of knowing whether (4.3.19) really represents a complex sentence or should rather be understood as direct speech:

(4.3.19) b. Then he said: "**another test** we will take"

In the remaining examples, the clause that is introduced by the topicalized constituent is independent. From the limited evidence I have, I would suggest that topicalization in embedded contexts as outlined by Bhatt is not a preferred option in spoken IndE; rather, topicalization tends to occur not only clause-initially, but also utterance-initially. Obviously, the absence of topicalization constructions in embedded contexts from the conversation files of ICE-India does not completely invalidate Bhatt's claim, which was made with respect to vernacular IndE to begin with.

The non-topicalized structure of examples (4.3.17) and (4.3.18) might look like the following, with *I think* and *it seems* as parentheticals rather than matrix clauses:

(4.3.17) c. She has **that**, I think
(4.3.18) c. She is getting **this**, it seems

Construed this way, the two examples would not necessarily constitute counterexamples to Bhatt's claim. A much larger corpus might be needed to find enough relevant discourse contexts for resorting to reported speech as in Bhatt's example,

which might then provide speakers with the necessary "wider range of grammatical contexts" for topicalizing. I do have my doubts, however, that complex structures such as the one quoted above ("*His friends know that her parents, he doesn't like at* all") represent naturally occurring instances of spoken **vernacular** IndE, but I am in no position to refute Bhatt's claim beyond the evidence that I have presented in this section.

4.3.3 Topicalization in ICE-India: Contexts and function

Many examples of topicalization in spoken IndE display the same function as in other varieties of English:[41]

(4.3.20) A: Hello <,> we'll arrange a party <,,>
B: Yes
A: On the fifth <?> who'll be happy </?> [uncertain transcription]
My birthday party <,,> you arrange [...]
Whereas **your birthday party** I'll arrange <,,>
(ICE-IND: S1A-003–6)

(4.3.21) A: Oh <,> we never knew what happened to her
The others <,> we knew <,> ah <,> but **Vidya** we didn't knew
(ICE-IND: S1A-021#58–59)

Examples such as these have been highlighted as illustrating the main function of topicalization, namely to indicate (some kind of) contrast:

> Object fronting is typically chosen when there is a communicative need to emphasize or contrast a clause element. Both the fronted element and the verb are strongly focused. (Biber et. al. 1999: 904)

This is intuitively obvious in both examples. In example (4.3.20), the fronted constituents ***my birthday party/your birthday party*** both belong to the subset of possible parties, which were explicitly mentioned in A's first utterance. Speaker A creates a contrast between the two upcoming birthday parties and their respective organizers. Similarly, in example (4.3.21), a contrast in terms of previous acquaintance is set up between ***the others***, who are already known, and one unknown person called ***Vidya***. However, Lambrecht has pointed out that "contrast" is not an inherent property or function of topicalization: "contrastiveness, unlike focus, is not a category of grammar but the result of general cognitive processes referred to as 'conversational implicatures'" (1994: 291). Example (4.3.22) shows

41. The overwhelming majority of examples in both ICE-GB and ICE-India are focus preposings.

that topicalization is not necessarily contrastive: A has started the conversation with *today I tell you* <,> *uhn* <,> *about the* <,,> *uh terrorism in Kashmir* and proceeds to give details about how her family was forced to leave the area:

(4.3.22) B: Then what happened <,,>
A: Yeah **something about Kashmir** I'll tell you because Kashmiri <,> in Kashmir Muslims are <,> also suffering a lot <,,>
(ICE-IND: S1A-054#34–35)

Birner & Ward attribute the observation that topicalization is not inherently contrastive to Prince (1984):

> contrast is not a *necessary* effect of topicalization, but "obtains just in case first, a list understanding is induced, and second, a salient opposition is inferred in the new information of the [open proposition]" ([Prince] 1984: 220). It does not follow from the salience of a set of entities that members of that set are being contrasted. Moreover, the entity represented by the preposed constituent in a topicalization is sometimes a singleton member of a poesetand [sic] thus need not be interpreted with respect to other discourse entities. (Birner & Ward 1998: 43)

Example (4.3.23) below provides the context for such a "list understanding": B is planning to send New Year cards to everyone and asks A for Sumi's address. During the course of the telephone conversation, B seem to realize that she has Sumi's address, after all, and that it is Meera's address she needs:

(4.3.23) B: I want Sumi's address *yaar*
I don't have her address
A: My God <,>
What a shame *yaar*
B: I <,> you tell me what to do <,>
A: Being her
B: *Haan nahi nahi* **Sumi's address** I have
A: Then <,>
B: I don't have Meera's address <,> (ICE-IND: S1A-098#78–86)

The opposition between *Sumi's address* and *Meera's address* is highly salient in the context. The following example satisfies just one of the discourse conditions which trigger a contrastive reading:

(4.3.24) So and we met Professor S T Naik <,> and **some other uh Naiks** we met <,> (ICE-IND: S1A-017#167:1:A)

Here, the "list understanding" concerns the set of people called Naik, but the relation between the two evoked members of the set – *Professor S T Naik and some other Naiks* – is additive, not contrastive.

If we now turn to further discourse functions of topicalization in spoken IndE, we find another function that is highly reminiscent of the function identified for a subset of the non-initial *there*-constructions treated in the preceding chapter. Many examples display an explicit discourse-linking function, where the immediately preceding topic is taken up again. This category includes examples such as (4.3.25–27) below as well as examples with preposed discourse-deictic anaphoric pronouns or NPs with a demonstrative determiner, as in (4.3.28–29) further below.

(4.3.25) A: In Jaipur then we have also <,,> we have a Birla *Mandir* and <,> of course we've a *Hawa Mahal* in Jaipur it's pretty famous <,> you must have seen photographs and all
C: Yes ***Hawa Mahal*** we heard (ICE-IND: S1A-008#57–58)[42]

(4.3.26) A: Studies <,,>
B: **Studies** I've not yet started <,,>
Studies <,> like I'd started just that day you told me you're studying *no* <,> (ICE-IND: S1A-040#215–217)

(4.3.27) A: Uhm uhm <,> what about M A <,> do you teach M A also?
B: Yeah I teach <,> uh literature at <,> uh the M A level
A: **Literature** you teach
B: Yeah <,> (ICE-IND: S1A-060#4–7)

In all the examples above, the main discourse motivation for preposing seems to be topic continuity. Example (4.3.26) is particularly interesting in that respect: A and B, two young female students, have been talking for some time when A changes the topic by introducing the topic of studies with a one-word utterance. *Studies* as the new topic is taken up in both of B's replies; in her first utterance ***Studies*** *I've not yet started* the NP *studies* is topicalized. In her second utterance, the NP *studies* is only loosely linked to the following proposition, indicated by the pause and the pragmatic marker *like*. Such "hanging topics" or "unlinked topics" will be discussed in Section 4.3.5.

Even more interesting are the examples with preposed anaphoric pronouns, most frequently discourse-deictic *that*. Table 4.19 shows that the proportions of preposed anaphoric pronouns is more or less the same in both corpora: 29 examples or 15.5% of all tokens in ICE-India and three examples or 15.79% of all tokens

42. (*Birla*) *Mandir*: Hindu temples built by the renowned business dynasty Birla. *Hawa Mahal:* the famous 'Palace of the Winds' in Jaipur, Rajasthan.

in ICE-GB fall into that category. These figures are well above those for Birner & Ward's corpus, where they found only 4% relevant examples (Birner & Ward 1998: 223). It has to be kept in mind, however, that their corpus is unbalanced and contains data from both spoken and written language. Gregory & Michaelis (2001: 1670) found 25% of all topicalization constructions in the Switchboard Corpus to contain a preposed anaphoric pronoun.[43] So, while the occurrence of anaphoric *that* as preposed constituent in spoken IndE is in principle not remarkable, the fact that such pronouns also occur in focus preposings is striking:

(4.3.28) C: But they have the other buses also *no* <,> then
A: Yeah
That you have <,>
B: Madras is known for that you know <,> TTB service **very best service** they have (ICE-IND: S1A-021#195–198)

(4.3.29) B: And Chicago linguistic society was the centre of learning
A: Yes **that** you mentioned <,> (ICE-IND: S1A-082#194–195)

In example (4.3.28), the three speakers reminisce about their student days, the conversation just centered on special bus services for students. When speaker A reacts to C's question with ***That*** *you have*, *that* must be the focus of the proposition, since the context does not licence any other interpretation: it is very difficult to imagine either *you* or *have* carrying another pitch accent. The same holds for example (4.3.29) because there is no discernible reason to assume that there is another prosodic peak in ***that*** *you mentioned*.

To be sure, not all examples with a preposed anaphoric pronoun are focus preposings:

(4.3.30) B: Particularly in Diwali days <,> they have to do *Laxmi pooja* and all that
A: Uhm there
B: So so really if you started seven till nine *pooja* and decoration and everything
A: **That** you must do previous day *na*
B: No **previous day** we never do (ICE-IND: S1A-065#188–192)[44]

43. It would be interesting to investigate whether the difference in frequency between ICE-India and ICE-GB on the one hand, and the Switchboard Corpus representing AmE, on the other hand, is an accidental by-product of the corpus designs or if it corresponds to a difference in discourse style in the respective varieties.

44. *Diwali*: Hindu festival of lights, celebrated in November; *Laxmi*: Hindu goddess of wealth, who is particularly worshipped during Diwali; *pooja*: ritual of worship.

When A tells B that ***That*** *you must do* PREVIOUS DAY, the focus is clearly on PREVIOUS DAY as the new information. Without access to the intonational contour of the actual utterance, B's reply *No* ***previous day*** *we never do* remains ambiguous: if the utterance carries another nuclear accent on *never*, then we might be dealing with another topicalization construction; if not, we have another instance of focus preposing where the preposed focus constituent forms an identity link to the prior discourse. The constraint proposed by Birner & Ward for focus preposings, namely that the relevant poset must contain at least one other member, does not hold for examples such as these: identity links are felicitous in spoken IndE in syntactic contexts in which they would not occur in spoken BrE or AmE. "Identity", then, can be translated as "topic continuity" in our context. Preposed anaphoric object pronouns serve as explicit links in the communicative situation, they are not required to represent new information in order to be felicitous.

Preposed anaphoric pronouns may also be combined with a quantifier and the universal quantifier may also occur on its own. In all these examples the superordinate term of the relevant poset is preposed: in example (4.3.31), speakers A and B discuss the necessary preparations for a marriage, and B cuts short A's questions with ***all that*** *we've done*.

(4.3.31) A: And have you booked the hall?
B: Yeah yeah that <,> **all that** we've done booked the hall <,> date etcetera (ICE-IND: S1A-095#171–172)

In the next dialogue about movies, speaker B again employs a superordinate term for the set of possible movies to watch:

(4.3.32) A: Have you seen *Evil Dead* <,>
B: Yeah <,> *Evil Dead* and
A: It's very nice
B: Very lovely film <,>
A: Then uh *Robo Cop* have you seen <,> *Robo Cop* and
B: **Those and all** I haven't seen <,> (ICE-IND: S1A-053#160–165)

In example (4.3.33), the universal quantifier *everything* effectively serves to round up all the previously mentioned topics.

(4.3.33) C: In fact when he is uh <,> dealing with the English literature he covers Sanskrit literature Kannada literature
A: Greek literature
C: Greek literature everything <,> British history **everything** he covers <,,> (ICE-IND: S1A-017#226–228)

This "summary" function is then very similar to the function identified in the preceding section for non-initial *there*-constructions with a discourse-old "notional" subject.

So far, the analysis has shown that one overarching discourse motivation for topicalization in spoken IndE is to express topic continuity. As said above, we owe the only in-depth studies on topicalization in other New Englishes to Mesthrie (1992, 1997). His analysis of topicalization in South African Indian English (SAIE) provides some interesting points of departure for comparison and further analysis.

Mesthrie notes that topicalization in SAIE appears in a much wider range of syntactic and discourse-pragmatic contexts than in L1 Englishes (1992: 113):

> SAIE goes beyond the functions associated with fronting and dislocation in mainstream varieties of English. Firstly, fronting may occur initially in a stretch of discourse, without any apparent recourse to givenness or contrast. [...] while such extratextual inferencing is possible in other English dialects, SAIE makes especially high use of it.

One of Mesthrie's examples illustrates his points. Upon arrival at his informants' homes, he was frequently asked:

(4.3.34) Your car where you parked? (= 'Where did you park your car?')
(1992: 114)

The utterance topic, namely his car, was "neither mentioned in the discourse not physically visible" (*ibid.*), that is, the topic was neither previously evoked in the discourse nor somehow 'given' in the speech situation, but was licensed by "extratextual inferencing". The same example may also be used to illustrate that topicalization "interacts to a much wider extent with other syntactic processes in SAIE than other English dialects" (*ibid*. 114). In this case, topicalization interacts with question formation.

My data contain only six examples of topicalization in interrogatives (two including *do* as operator). All are focus preposings where the focus referent is cognitively accessible to the listener: in example (4.3.35), the conversation between three female teachers has been about husbands and children, so that a question about other relatives does not come as a surprise.

(4.3.35) C: **Brothers and sisters** you have <,>
A: Yes <,> I have two brothers and two sisters
C: Oh all are in India <,>
A: All are in Madras <,> except my elder sister
She is in Pondecherry <,> (ICE-IND: S1A-030#226–230)

The communicative setting of the dialogue from which example (4.3.36) is taken is similar: a teacher from Rajasthan and a teacher from Goa are comparing the festivals and holidays that are being celebrated in their respective states. When speaker B asks about *Holi*, one of the most prominent festivals across India, the question has a discourse-new topic which is nevertheless hearer-old.

(4.3.36) B: **Holi** do you have
A: Yeah **Holi** also we have (ICE-IND: S1A-065#136–137)[45]

Example (4.3.37) again conforms to the pattern that no "extratextual inferencing" is involved in IndE topicalized interrogatives because the topic of admission to colleges and degree courses has already been evoked in this conversation among students.

(4.3.37) B: Leena Raina <,,> Laxmi all are joining engineering
A: No one is joining medicine <,>
B: No <,> nobody is interested because
A: **Admissions** you are getting <,>
B: Yeah they'll easily get <,> They are not interested
(ICE-IND: S1A-040#236–241)

Two of the interrogatives can be described as "echoings", a subtype of focus preposings:

> As with focus preposing in general, the link and focus of echoing are one and the same; however, unlike other types of focus preposing, the link of echoing typically represents a poset member that has been explicitly evoked in the prior discourse. Preposings of this type are used by speakers to convey their lack of commitment as to whether the link constitutes an appropriate instantiation of the focus variable within the open proposition. A speaker may wish to convey such a lack of commitment either because he or she is uncertain about the appropriateness of the link or because he or she is doubtful about it. (Birner & Ward 1998: 88)

Birner & Ward provide the following example: a Burger King employee asks the customer *Large Coke you ordered?* (*ibid*.) about five minutes after the customer has placed the order, obviously because he or she was not quite sure to have taken the order correctly. In both examples below, the communicative situation is similar. In example (4.3.38), three college lecturers are talking about their jobs. Speaker B seems to express her incredulity that C did not have to do 'evaluation':

45. *Holi*: a festival celebrated in early march which involves dousing people with colours ('playing Holi').

(4.3.38) A: As soon as the exams were over <,> evaluation <,,> and then <,> uh I just finished evaluation and came here <,>
C: I didn't go anywhere *no* <,,>
Just I <,> kill the time in summer <,> reading <,> doing nothing
B: **Evaluation** you have done <,>? (ICE-IND: S1A-031#296–299)

The telephone conversation in (4.3.39) turned around an order of tin oxide and the question of whether ordering 50kg or 100kg might be cheaper in terms of air-freight charges. Speaker A asks once more *but **hundred k g's** you don't want?* to be absolutely sure that he understood his business partner correctly:

(4.3.39) B: Yeah <,> you ask them to go ahead
A: *Haan* <,> but **hundred k g's** you don't want <,>?
B: No hundred k g's
A: **But fifty k g's** there might be some price implications I'm trying to verify why <,> (ICE-IND: S1A-094#123–126)

Speaker A's second turn, ***But fifty k g's** there might be some price implications*, involves another 'hanging topic', to which I will return below.

A minor point, which nevertheless deserves to be made, concerns the topicalization of adjectival subject complements (AP topicalizations): the interesting thing here is that the findings from both ICE-GB and ICE-India do not corroborate Birner & Ward's analysis at all. AP topicalization is rare in their corpus (17 out of 529 topicalization tokens, i.e. 3.21%), whereas the figures for both ICE-India and ICE-GB are much higher: 10.7% and 10.52% of all relevant tokens are adjectival subject complements. According to Birner & Ward, "the type of topicalization permitting preposed APs typically involves a salient set of two attributes, one of which is affirmed, the other denied, used in tandem to *evaluate* some salient discourse entity" (1998: 47). A prototypical example for them would then be the following:

(4.3.40) N: This is not another vulgar disgusting sexploitation film.
J: *Vulgar it's not. Dumb it is.* Did we see the same movie? (1998: 47)

However, with "single AP topicalizations, all of the posetmembers must be explicitly evoked" (1998: 44):

(4.3.41) G: I can't stand R. He's stupid, arrogant and totally off-the-wall.
B: *Stupid I wouldn't really say he is.* (1998: 44)

They conclude: "it appears that AP topicalization is restricted to the explicit contrast of one anchoring poset member with another (not all of which need appear in preposed position)" (1998: 48). However, neither the ICE-GB nor the ICE-India examples seem to conform to the scenario identified as the most typical for

preposed APs. The two examples from ICE-GB do not express contrast: in example (4.3.42), speaker E complains about her cat leaving through the cat flap every morning to a sympathetic audience, and in example (4.3.43), the speaker expresses her gratitude to B. Even though two examples are not enough to attempt alternative generalizations, we may remark that in both cases, the topicalized adjective appears to summarize and capture the gist of the immediately preceding conversational exchange. In that sense, the information conveyed by the preposed constituent is discourse-new, but highly predictable:

(4.3.42) E: She's like an alarm clock
She wakes me up at six every morning
F: Mm
C: Yeah
E: **Horrible** it is (ICE-GB: S1A-019 #199–203)

(4.3.43) C: Have you got his album
B: Yeah
C: I'd really love to tape it from you if you if you didn't mind
B: Yeah
If you give me a tape I've got a tape to tape and I can run it off
C: Oh **great** that'd be (ICE-GB: S1A-042 #127–132)

The 20 examples from ICE-India do not lend themselves to a unified analysis, either, except that they also fail to conform to the constraints that derive from Birner & Ward's analysis. The only example where some kind of contrast is evoked is the following:

(4.3.44) A: Mean time I request you to come this side with <,> the photograph album
B: Yeah <,>
A: If possible
B: Yeah <,>
A: And show it to my wife <,> who is very keen <,>
B: Okay
A: Of course I am also keen
B: Ah
A: But **more keen** she is <,> (ICE-IND: S1A-095#366–374)

A group of examples occurs as immediate response to a question, e.g.:

(4.3.45) A: How is <,,> Akshay?
B: **Nice** he is
A: Ahn (ICE-IND: S1A-093#85–87)

(4.3.46) B: Then uh <,> then uhm how are your colleagues in office <,>
A: **Okay okay** they are <,> (ICE-IND: S1A-049#181–182)

The preposed predicative adjectives here necessarily represent new information in interrogative contexts, but in all other examples, the information status of the preposed constituent is easily accessible to the listener by virtue of being linked to the preceding discourse.

In example (4.3.47), the speaker talks about the Indian novelist Premchand (1880–1936) and his depiction of the hardships of rural life:

(4.3.47) A: Poverty is lying there all glaring poverty under the repressive rules of taxes
All type of taxes being levied by the aristocratic and ruling set up
These are all glaring and all visible <,> **clearly visible** it is I think so
(ICE-IND: S1A-006#37–39)

The excerpt in (4.3.48) comes from speaker B's report about a trip to town, and the last example (4.3.49) comes from a discussion about the practice of "ragging" among students, i.e. some kind of test of courage inflicted upon first year students. In this particular case, a student was forced to eat a bar of soap:

(4.3.48) B: [...] and then we went there and we bought cloth and all and we've gone to eat something you know both of us were sleeping there in the restaurant
So tired we were <,> we couldn't talk also anything but <,> while coming home it was quite <,> nice you know the journey
(ICE-IND: S1A-040#70–71)

(4.3.49) B: So by the time he finished the soap he was <,> admitted in the hospital <,>
A: *Accha*
B: **That <,> that serious** he became <,,>
(ICE-IND: S1A-090#112–114)

The last two examples are similar to constructions such as *Very interesting it is*, a "construction that is gaining ground in highly informal SAIE" (Meshtrie 1992: 114) with an intensifier such as *so*, *too*, or *very* preceding the adjectival complement. I do not have enough examples to assign this construction to a highly informal register of IndE, but its occurrence in a corpus of standard spoken IndE suggests its acceptability in neutral contexts.

The data from both corpora then provide clear evidence that AP topicalization is not as highly constrained as Birner & Ward assume. The kind of examples which inform their analysis are completely absent from my data, and with the exception

of AP topicalization in response to questions in IndE, preposed predicative adjectives do not behave any differently from other preposed constituents in IndE.

4.3.4 Preposing of non-arguments

As said above, the focus of this chapter is on preposing in the narrow sense, i.e. I exclude preposed adjuncts, as the line between marked and unmarked constructions is very difficult to draw indeed. I therefore limit myself to some tentative observations on the range of preposing options in spoken IndE.

It seems that preposing of adverbs is not restricted to sentence adverbs in spoken IndE, e.g.

(4.3.50) A: You have not seen even Mysore's <,> marvellous palace <,>
B: Sometimes I come along with my <,> students
A: **Previously** you had visited
B: Yes yes once <,> I had been to Palace <,,>
(ICE-IND: S1A-022#118–121)

(4.3.51) Grades are fixed <,> as soon as you get appointment <,,> uh uh when the document will be given to you <,> and in that **clearly** it is mentioned <,,> in which grade you are <,> (ICE-IND: S1A-063#259:1:B)[46]

(4.3.52) But their parents whenever I beat them <,> the parents came and ask me <,> purposely why <,>
And the same parents said **unnecessarily** you beat my child
(ICE-IND: S1A-085#167–168)[47]

(4.3.53) B: So now you see <,> they have to immediately ship <,> fifty k g <,>
A: **Immediately** they say it's not possible (ICE-IND: S1A-094#81–82)

To the best of my knowledge, preposing of adverbs other than sentential adverbs is not attested for any other variety of English. The range of adverbials that may appear in prenuclear position also seems less restricted than in ICE-GB, e.g.:

(4.3.54) A: So then you are coming to Goa <,>
B: [One word] Yes shortly
A: Let me know now
B: **With my family** I'll come (ICE-IND: S1A-001#66–70)

46. *Grade* refers to the grade/level on the salary scale for a particular job.

47. Speaker A, a female high school teacher, is complaining to a colleague that the parents of bad pupils never come to the school to talk to her, except to complain when she beats their children.

(4.3.55) A: That friend of mine she <,> can put on her own hand <,> she can put by left hand also <,> on her right hand
B: Oh like <,> uh no she can't do like that <,,>
A: **And for putting on other people's hand** <,> she is expert <,>
(ICE-IND: S1A-044#183185)[48]

Again, as I have no clear-cut criteria for including or excluding individual tokens as particularly marked examples of preposing, I will not pursue this matter further. Mesthrie notes for SAIE that it "is less common to form topics from oblique NPs, apart from temporal NPs and locatives" (1992: 116), the examples above, however, feature preposed constituents which go beyond the semantic roles mentioned by Mesthrie.

What is worth noting is that the ICE-India data feature more than 30 examples of preposed adverbial PPs where the preposition is deleted, e.g.:

(4.3.56) A: By six I will be at Shrinath's house <,>
B: Still it'll be late because **Nipani** we don't get bus back to Sankeshwar
(ICE-IND: S1A-058#233–234)[49]

The speakers are discussing a visit to their friend's house in Nipani, and speaker B remarks that it will take them a while to get back home because they will not get a bus back [***from***] ***Nipani*** to Sankeshwar. Mesthrie makes an intriguing comment concerning similar constructions in SAIE: he considers some examples of fronted locatives as an "example of promotion of a locative to the status of 'proper' topic [... which] involves the optional deletion of the preposition of the fronted PP", e.g.:

(4.3.57) Temple – they poison your mind. (= 'They poison your mind *in* a temple') (Mesthrie 1992: 117)

According to Mesthrie. the NP *temple* here becomes the utterance topic by virtue of being placed in prenuclear position and by deletion of the preposition. However, without further context, it is difficult to evaluate whether *temple* in this example really is a topic in the sense of Lambrecht (1994: 131):

> Topic: A referent is interpreted as the topic of a proposition if in a given discourse the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent.
> Topic expression: A constituent is a topic expression if the proposition expressed by the clause with which it is associated is pragmatically construed as being about the referent of this constituent.

48. Speaker A's friend who features in this extract is an expert in applying *mehndi* (Henna bodypaintings).

49. *Nipani*: town in Karnataka, South India; *Sankeshwar*: another smaller town close to Nipani.

With this definition, Lambrecht has made the intuition more precise that the topic is what the utterance is "about". Similar examples from ICE-India are again not restricted to locatives, and here we have the opportunity to consider the wider discourse context in order to evaluate whether the preposed PP with deleted preposition is indeed the topic. In example (4.3.58), two employees of an international company are having a telephone conversation about an internal problem, and a person named R. K. Mehta (who seems to be a senior member of the company) is repeatedly mentioned as the one who might be able to solve it. The utterance *Actually* [*on*] ***Sunday*** *they could talk to him* is not 'about' *Sunday* in any sense, but about the aforementioned R. K. Mehta: [*on*] *Sunday* represents new information and is thus more akin to a focus than a topic expression (if we disregard its categorical status as an adjunct).

(4.3.58) A: *Accha* <,> today afternoon R K Mehta's call was there <,> and he was again reminding on this uh <,>
B: Ah
B: You told him *na* but <,> problem <,> you told him the problem
A: Ahn
Uh
Actually **Sunday** they could talk to him <,> he has already updated <,> <,> so <,> now <,,> he is coming on Sunday
(ICE-IND: S1A-094#232–237)

Similarly, the temporal expressions [*on the*] *tenth,* [*on the*] *twenty-first,* [*on the*] *twenty-ninth* etc. in example (4.3.59) are not topics, but information that is predicated about the topic, namely the speaker and his family:

(4.3.59) A Uhm I mean **tenth** I'm leaving
B *Accha*
A With my family to Bangalore
B Yes
A Then <,> **twenty-first** we'll leave for Bombay <,,>
Twenty-ninth we'll start from Bombay in <.> fi </.> **thirtieth** I'll reach <,> (ICE-IND: S1A-019#64–69)

Another set of examples with fronted PPs that are realized with the preposition omitted occurs in response to questions, e.g.:

(4.3.60) A: And how long were you <,> alone in Poona <,>?
C: **Poona** I would have been alone for say <,>
A: Two three months <,>
C: No <,> one month (ICE-IND: S1A-099#304–307)

(4.3.61) B: You don't know anything about craft aunty <,>
A: **Craft** I don't know much <,>
B: Because I wanted to learn something <,> uh which you know <,,>
(ICE-IND: S1A-044#81–83)

In both examples, the speakers are reacting to a question that contained the full PP and begin their answer with a repetition of the noun, omitting the preposition. Again, these nouns are not topics because the topic expression in the proposition ***Poona** I would have been alone for say*... is 'I', referring to the speaker. The discourse motivation for this strategy of explicitly repeating the final element of the preceding speaker's turn – reduced to or recategorized as NP rather than PP – once more seems to be topic continuity.[50] In a further set of examples, the same desire for topic continuity seems to prompt the speakers to create a link to their own proposition; as in speaker A's final turn in example (4.3.62) below and speaker B's statement, repeating himself by uttering ***South** how lenient they are*. I would intuitively classify all three preposings as adjuncts which supply additional (locative, temporal etc.) information, and not as topics.

(4.3.62) A: And what about monsoon
B: **Monsoon** we rarely get any rains <,> uh say say in September <,> when the monsoon is receding from India then we get rains for four or five days and that too very slightly
A: So nice you know like that <,> hope it was the case in Goa also
B: You have heavy rains
A: Yeah **Goa** it's raining very heavily and you can't go anywhere out also <,> uhm (ICE-IND: S1A-065#16–20)

(4.3.63) B: So <,,> those are the differ[e]nces you can find like how <,> badly they are ragging in north and <,> south <,,>
South how lenient they are <,,> (ICE-IND: S1A-090#129–130)

4.3.5 Unlinked topic constructions

Lambrecht refers to a construction which he labels "unlinked topic construction", which "involves detached lexical noun phrases which have no anaphoric link with a pronominal topic expression inside the clause" (1994: 193) and which are restricted to informal spoken language. e.g.:

50. In example (4.3.61), speaker B's utterance *You don't know anything about craft aunty* ends with the vocative *auntie*, which plays no role for A's following turn in establishing topic continuity.

(4.3.64) (Talking about how to grow flowers) Tulips, you have to plant new bulbs every year? (*ibid.*)

Other scholars who have dealt with this construction have proposed alternative designations such as "hanging-topic constructions" (Primus 2001: 884). Gundel treats examples such as *My work, I'm going crazy* (1988: 224) as a subtype of left dislocation and states: "Such sentences, referred to (somewhat misleadingly) as 'double-subject' constructions, are topic-comment structures par excellence" (*ibid.*). In more formal English, the constructions is possible as *As for my work, I'm going crazy*, with *as for* serving as a topic-introducing device. Without *as for* preceding the unlinked topic, the construction would in formal English be frowned upon as the stylistic vice of anacoluthon, defined by the OED as a "want of grammatical sequence; the passing from one construction to another before the former is completed".

Lambrecht shares Gundel's categorisation of the construction as being closer in function to LD than to topicalization constructions: "the *as-for* construction is [...] a subtype of the detachment or dislocation construction" (1994: 152). Left and right dislocation will be the topic of the next chapter, yet, for the time being, it is interesting to note that topicalization and left dislocation may not be as categorically distinct as Ward & Birner presume.

The ICE-GB conversation files contain only three relevant examples of unlinked topics:[51]

(4.3.65) **Rugby** <,,> the girls are just treated like a few honorary girls but they're not integrated (ICE-GB: S1A-054#100:1:A)

(4.3.66) I mean **basically U C** if you come here they give everybody a piece of paper all postgraduates a piece of paper saying there is no funding I'm afraid [laughter] (ICE-GB: S1A-079#054:1:A)[52]

(4.3.67) But you know **a few people in this department** I definitely find my style of speech changes dramatically because they use a <,> more extensive vocabulary (ICE-GB: S1A-084#083:1:B)

Lambrecht notes that such constructions are typical of topic-prominent languages (2001b: 1058, cf. Chapter 3.2.3 above). They obviously do occur as well in subject-prominent languages such as English, but are restricted to informal conversation. With only five examples, the ICE-India direct conversation files do not yield a substantially higher number of unlinked topic constructions than ICE-GB:

51. The examples were retrieved when searching for left and right dislocation (included in the function (DEFUNC,NP)).

52. *UC*: University College (London).

(4.3.68) A: Good going <,> I think we'll have a change in the topic now <,> uh we will talk of the examination system <,,> uh to which all of us are agreed to <,,>
Yes why don't you begin <,,>
D: **Exam system** what do you feel that <,,> uh you are getting your rights <,,> the exam present kind of exams
C: No ma'am I feel <,> examination should be abolished <,>
(ICE-IND: S1A-011#99–102)

(4.3.69) A: So when you write to your parents which language you write in <,,>
B: To my father English <,> to my mother <,,> Telugu <,> to my grandparents Telugu <,> aunt Telugu
A: So you know the written variety of Telugu also <,>
B: **The written varieties** I write down what I speak <,> that should not be [one word] <,,> (ICE-IND: S1A-014#154–157)

(4.3.70) A: And what about you like <,> you are doing a computer course now-a-days
B: Yeah <,> right now I am doing my computer course so <,> so after the MA I think I better do my computers <,> continue my computers <,>
A: Continue your computers and what about the job <,> or something
B: **Job** let's see <,> (ICE-IND: S1A-062#162–165)

(4.3.71) B: I am scared about linguistics <,> because we have to be very specific while writing the <,> answers we have to study the definitions and all that <,>
A: And I think grammar is worst <,>
B: But **grammar** we can apply what we have studied to something
(ICE-IND: S1A-064#9–11)

(4.3.72) B: You can read Hindi <,>
A: Oh yes sir
B: Means you can read Marathi also
A: Yes Hindi I can read Hindi also
B: So Marathi and Hindi <,>
A: Oh yes **script** they are the same <,> I can read Marathi without any problem (ICE-IND: S1A-076#126)

In examples (4.3.69), (4.3.70) and (4.3.71) above, the unlinked topic has just been explicitly mentioned in the preceding utterance, following the pattern of establishing topic continuity by repetition that has already been mentioned.

To summarize: the simplest observation on topicalization in spoken IndE concerns its frequency: topicalization is indeed much more common in IndE than in BrE, which might be taken to indicate that it is also less marked as a construction. The evidence for the construction's extension to more syntactic environments, as claimed by Mesthrie & Bhatt (2008: 81), remains ambiguous: from the few relevant examples available, we cannot say for sure whether topicalization regularly occurs in embedded contexts in spoken IndE. On the other hand, our examples of topicalization in interrogatives support Mesthrie's analysis based on SAIE (1992: 114) that topicalization may interact with a wider range of syntactic processes in New Englishes. Mesthrie has further found that topics in SAIE may be licensed by "extratextual inferencing", that is, topicalized constituents are not required to establish a discourse-old link to what has been said before and thus violate Ward & Birner's main criterion for felicitous preposings. However, none of the ICE-India examples fails to adhere to this requirement. Rather on the contrary, it is even extended to contexts where it does not hold in BrE or AmE, namely focus preposing involving "identity links". An IndE topicalization construction is thus much less likely to favour a contrastive reading, and much more likely to promote overt topic continuity in discourse.

4.3.6 Explanatory parameters

Turning first to the correlation of topicalization constructions with speaker variables, we find the parameters gender, age and genetic affiliation to be highly significant (at $p = <0.0001$). Women are overrepresented with respect to the constructions, as are the members of the two youngest age groups in the ICE-India speaker sample.

The variable "educational level" is not statistically significant ($p = 0.2982$). In particular, we do not find the construction to be markedly underrepresented with those who hold a PhD.

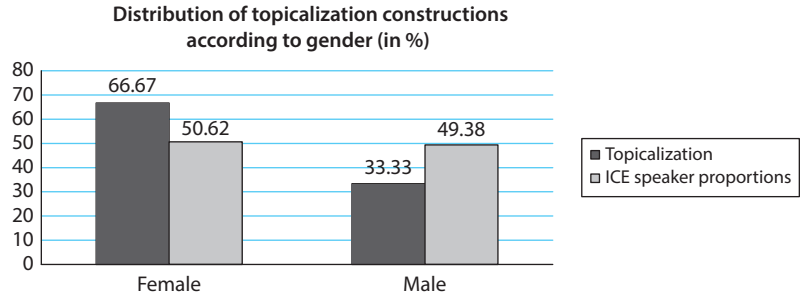

Figure 4.20. Distribution of topicalization constructions in the ICE-India conversation files according to speaker's gender

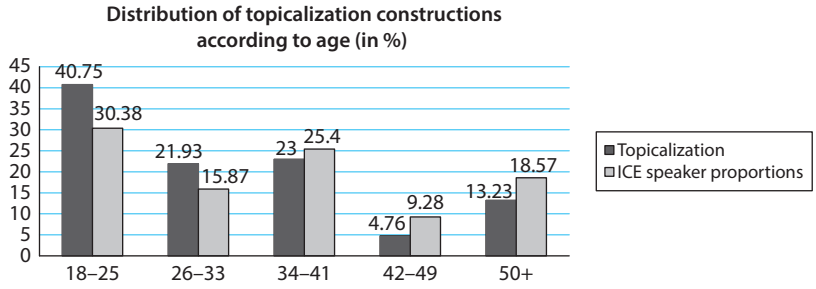

Figure 4.21. Distribution of topicalization constructions in the ICE-India conversation files according to speaker's age

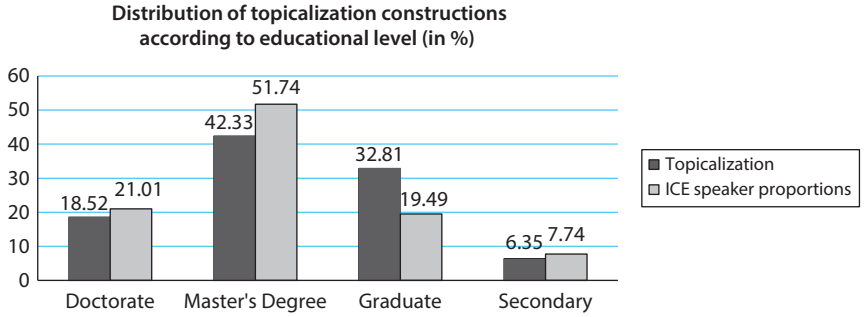

Figure 4.22. Distribution of topicalization constructions in the ICE-India conversation files according to speaker's educational level

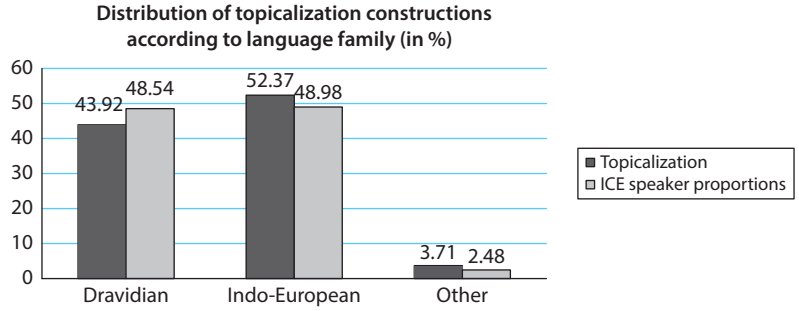

Figure 4.23. Distribution of topicalization constructions in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

The correlation with language family is again significant ($p = <0.0001$), but there is no drastic imbalance between speakers of the Dravidian and the Indo-European languages.

The causes or motivations for the higher incidence of topicalizations in spoken IndE compared to spoken BrE, or in New Englishes generally, are notoriously difficult to pin down. Since the phenomenon appears to be pervasive across many Asian and African Englishes, to look for substrate influence seems pointless. Mesthrie, who has analysed topicalization in SAIE "as illustrative of the process of building syntax out of discourse in the process of language shift" (1992: 110), has commented on the feasibility of substrate influence for this particular phenomenon (1992: 157):

> The predilection for topicalisation [...] is not substrate-induced. The Indic and Dravidian languages do not appear to use a particularly striking proportion of topicalised sentences (no more than standard English, say). Once again we see universals of discourse structure playing a greater role than transfer.

Mesthrie refers to Givón's programmatic book *On Understanding Grammar* (1979b) to elucidate the notion of 'universals of discourse structure':

> Givón's distinction (1979: 223) between the pragmatic and syntactic mode of communication would seem to be pertinent to SAIE, Singapore English and New Englishes generally. [...] The high use of topic-comment structures would seem to be characteristic of language-learning situations – pidgins, child language and adult second-language acquisition. Although use of the strategy might be reduced in formal discourse, it is manifested in informal speech. (Mesthrie 1992: 123f.)

Several problems arise if we want to apply this perspective to our corpus findings. Firstly, neither SAIE nor Singapore English are directly comparable with spoken IndE as sampled in ICE-India. SAIE as described by Mesthrie is a language-shift variety, with speakers representing all stages from pre-basilectal to acrolectal proficiency. Colloquial Singapore English or Singlish, in turn, may be more advanced than IndE in terms of Schneider's Dynamic Model (cf. Mukherjee & Gries 2009), but then the communicative space occupied by English in Singapore is highly diglossic, with standard (exonormative) English as the H-variety for writing and Singlish as the popular everyday L-variety. An argument which falls back on universals of language learning seems to me difficult to sustain when applied to the speakers represented in ICE-India. After all, Mesthrie's conclusion that SAIE "does seem to have the properties of topic-prominent languages" (Mesthrie 1992: 124) which prompts him to place the variety within the realm of the pragmatic mode of communication explicitly rests on the syntax of *basilectal* SAIE.

The second problem is that the equation of topicalization with topic-comment structure, as in the quote above, or with a general "predilection for topics" (*ibid.*) is largely taken for granted. We have seen above that preposing may operate either on topics or on foci, and Gundel & Fretheim confirm that "non-canonical placement of constituents in sentence-initial position is not in itself uniquely associated with either topic or focus" (2004: 185). It is this absence of a one-to-one

correspondence between syntactic configuration and discourse-pragmatic function which makes the analysis of topicalization so difficult, and which also prohibits the sweeping generalization that everything which ends up in prenuclear position is a topic. This latter point is particularly evident for non-argument preposings: Mesthrie's notion of topicalization includes fronted locative and temporal PPs as well, and he has analysed the fronting of PPs together with the deletion of the preposition as "promotion to 'proper' topichood" (cf. above). However, I have shown above that the preposed constituents in parallel examples from ICE-India are not topics. I would further expect more unlinked topic constructions in spoken IndE if topic-prominence as an instantiation of the pragmatic mode of communication were the main explanatory parameter in this context.

Looking at all the evidence available, I would strongly suggest to take a closer look at substrate influence as the decisive factor for the form and frequency of topicalization constructions in spoken IndE, contrary to Mesthrie's rejection of this element. The evidence that speakers make use of patterns familiar from other languages which belong to their overall linguistic repertoire comes from two different observations.

Firstly, I noted that the syntactic contexts for topicalization in IndE are much less restricted compared to BrE or AmE. This comparative syntactic freedom is also apparent in the frequent preposing of non-arguments (cf. Section 4.3.4). I would still hesitate to extend the notion of "topic" to every constituent that is moved to sentence-initial position, but what we see in our examples are clear cases of pattern replication. In Chapter 3.2.3 above, I referred to some structural properties shared by all Indian languages and noted that flexibility of word order is a pan-Indian feature. In particular, all Indian languages allow movement of more or less any constituent to sentence-initial position. This freedom from syntactic restrictions on the type of constituent that is topicalized is then transferred to IndE. To put it the other way round, speakers of IndE map the syntax of Indian topicalization constructions onto English, which results in the loss of syntactic restrictions on the English construction. Pattern replication typically affects syntactic and selectional restrictions of a construction in the actual process of replication. When it comes to the syntax of information structure, such a shift in the syntactic profile of a construction will also entail a shift in its discourse-pragmatic meaning. When, as seems to be happening in IndE, the syntax of topicalization is extended to contexts beyond its original definition, then we might be witnessing the process of unmarking a marked word order pattern:

> a bilingual under the influence of the norms of a second language may start using a marked [...] alternative order in his own language with greater frequency than is normal until finally it starts functioning as an unmarked order for him and others like him, whom monolinguals in turn imitate, with the change – a shift in stylistic status – so gradual that it is hardly noticed. (Masica 1976: 15f.)

We will come back to this "loss of pragmatic specialisation of secondary word order variants" (Matras 2009: 251) in subsequent chapters.

The second observation that I want to enlist for my argument concerns the discourse contexts for IndE topicalization constructions. IndE speakers have extended the range of discourse functions of topicalization to include the function of elliptical repetition (cf. 4.2.6.1 above). To explicitly refer to or to repeat part of what the previous speaker has just said creates cohesion in discourse as well as among the participants in a conversation. This function is particularly apparent where it overrides the discourse constraints on topicalization, for example when it allows identity links in focus preposings as shown above. Thus, topicalization in IndE in this particular function is an instantiation of the pan-Indian "grammar of culture" (cf. Section 3.2.4), again a substrate feature. The question that presents itself then is whether the process of pattern replication and the process of transferring pragmatic patterns are related. I will reserve the discussion of this question for my final chapter.

4.4 Dislocation

Platt et al. (1984: 119–121) have listed syntactic constructions which involve the addition of a resumptive pronoun as very common across all New Englishes. They provide examples for "pronoun copying [...] the practice of adding a pronoun after the noun subject of a sentence" (1984: 119), i.e. left dislocation (LD) as in:

(4.4.1) You see there's a terrible there's a terrible feeling I mean nervous young singers **they** don't want to look at the audience <,>
(ICE-GB: S1A-045#090:1:B)

(4.4.2) Yes clapping and singing he's against **it** three hours non-stop [laugh]
(ICE-GB: S1A-068#251:1:C)

Example (4.4.1) represents the prototypical case of a dislocated subject, example (4.4.2) shows that dislocation is not confined to subjects. According to Platt et al. (1984: 120),

> This occurs in the 'older' Englishes as well but here it is confined, strangely enough, to a rather pompous oratorical style or a very colloquial style [...] it is a particularly useful device for speakers of the New Englishes who do not make the same use of intonation for emphasizing as do some of the speakers of British English.

However, the *Handbook of Varieties of English* (Kortmann et al. 2004) does not follow up this lead. The only related morphosyntactic feature that finds recognition in the handbook's global synopsis is feature 67 "resumptive/shadow pronouns

(e.g. *This is the house which I painted it yesterday*)" (Kortmann & Szmrecsanyi 2004: 1148), listed under "relativization". This feature is also top on the list of morphosyntactic phenomena that are the exclusive property of L2 Englishes in Asia, the Pacific and Africa (2004: 1192). However, while many scholars refer to the prevalence of dislocation in African and Asian Englishes (cf. the contributions in Mesthrie 2008), the relativization strategy involving resumptive pronouns that is singled out for the *Handbook* seems to be quite rare in these varieties. Mesthrie (1992: 79) provides some examples from SAIE and labels the construction "a rare discourse-governed strategy which has not stabilised". ICE-India features only eight examples from seven speakers, e.g.:

(4.4.3) And asthama is such a <,,> disease that **it** can be completely cured by homeopathy only (ICE-IND: S1A-037#17:1:A)[53]

It seems safe to assume that resumptive pronouns in relative clauses are much more marginal than suggested in the *Handbook*. To be sure, the presence of resumptive pronouns is a necessary, but not a sufficient condition for a construction to be labelled "dislocation" or "detachment construction", as the following section will clarify. I will then discuss the suggestions that are available in the literature concerning the discourse function of dislocation constructions in contrast to other topic-marking devices such as topicalization and existential constructions (cf. Chapters 4.2 and 4.3). Sections 4.4.4 and 4.4.5 then present the relevant data on form, function and frequency of both types of dislocation from ICE-GB and ICE-India. By now, it will not come as a complete surprise that the type of dislocation singled out by Platt et al., namely left dislocation, occurs with a considerably higher frequency in the ICE-India data than in the corresponding data from ICE-GB. In a final section, I will discuss possible motivations for the prevalence of the construction in spoken IndE. Note that Platt et al. have already hinted at an explanation for LD in the New Englishes generally: the construction with a resumptive pronoun is considered to be equivalent to the use of intonation for the purpose of marking emphasis. I will return to their suggestion below.

4.4.1 Definition

Lambrecht proposes the following formal definition of dislocation constructions (2001b: 1050):

> A dislocation construction (also called detachment construction) is a sentence structure in which a referential constituent which could function as an argument or adjunct within a predicate-argument structure occurs instead outside the

53. *Asthama*: asthma; speaker A is arguing for homeopathic rather than allopathic treatment.

> boundaries of the clause containing the predicate, either to its left (left-dislocation, henceforth LD) or to its right (right-dislocation, henceforth RD). The role of the denotatum of the dislocated constituent as an argument or adjunct of the predicate is represented within the clause by a pronominal element which is construed as coreferential with the dislocated phrase. Typically, the dislocated phrase is marked with special prosodic features.

This definition applies to left dislocation as in examples (4.4.1) and (4.4.2) above as well as to right dislocation (RD), e.g.:

(4.4.4) **He's** a funny chap that uh solicitor [unclear-words]
(ICE-GB: S1A-061#185:1:A)

(4.4.5) Yeah i i i **it** is very odd this whole aspect of code switching
(ICE-GB: S1A-056#040:1:A)

The definition excludes relativization involving resumptive pronouns as in example (4.4.3) above because no referential constituent occurs outside of the clause boundary. I will therefore restrict the discussion to single main clauses with a left- or right-dislocated constituent and a coreferential resumptive pronoun.

4.4.2 Left dislocation: Form and function

The surface structure of left dislocation resembles topicalization: in both constructions, a constituent is preposed to sentence-initial position. The defining syntactic criterion to distinguish the two constructions is the presence of a coreferential pronoun occupying the syntactic slot which the preposed constituent would occupy in the canonical form of the sentence. This syntactic configuration also allows subjects to be left-dislocated (cf. example (4.4.1) above), which, in their turn, cannot be topicalized. However, "the formal distinction between the two types of construction corresponds to a functional distinction", as Birner & Ward (1998: 94) have noted: while the preposed constituent in topicalization typically establishes a discourse-old link to the preceding discourse, there are no such restrictions on the information status of a dislocated constituent.

Several researchers have tried to establish a unitary function for left dislocation. For Gundel (1988: 211), resumptive pronouns in LD constructions may serve to unambiguously mark the topic, as in her example *Marcos, **he** resigned* where the pronoun overtly identifies *Marcos* as the syntactic topic. A further related function for LD constructions arises when topics are being reintroduced by speakers into the discourse, e.g.:

(4.4.6) That plant you gave me, it's really grown. (Gundel 1988: 212)

The topic expression *that plant you gave me* may refer to an entity that is not currently the centre of attention for the participants in the conversation. The LD construction then serves to retrieve the topic as salient in the current speech situation. It is this function of LD that Givón has singled out (1993: 210, emphasis in the original):

> Left dislocation is used to mark important referents – most commonly definite or generic – that are brought back into the discourse after a considerable **gap of absence**. That such referents are anaphorically accessible is suggested by the fact that they may be either definite or generic, but never referring-indefinite. In other words, L-dislocation is not used to introduce new referents into the discourse.

Prince, however, argues that "no single function can in fact account for all the Left-Dislocation data in English" (1997: 120). Prince distinguishes three types of LD constructions, of which two are relevant in this context.[54] The first type is called "simplifying LD" because of their function within discourse processing:

> A 'simplifying' Left-Dislocation serves to simplify the discourse processing of Discourse-new entities by removing them from a syntactic position disfavored for Discourse-new entities and creating a separate processing unit for them. Once that unit is processed and they have become Discourse-old, they may comfortably occur in their positions within the clause as pronouns. (Prince 1997: 124)

The following excerpt from a conversation may serve as an example: speaker B tells A about her experience as a member of the "Kingston Conservation Volunteers" and how she got involved with the group:

(4.4.7) B: Yeah we go back <,> you know when we're needed you know
My friend Alice who lives opposite <,> **she** organises the task
She rings up the wardens <,> and says you know
A: Yeah
So you got involved through her did you
B: No I she got involved through me
A: Ah <„>
B: My ex-boyfriend Phil **he** got me interested
I went out <,> one
It was December the thirty-first nineteen eighty-nine and I really enjoyed it and then Alice came out on my second task and she's loved it (ICE-GB: S1A-081#117–127)

Speaker A resorts twice to left dislocation to introduce discourse-new referents in her narrative, first *my friend Alice* and then *my ex-boyfriend Phil*. Once the two

54. Her third type of LD concerns resumptive pronouns in relative clauses, which are excluded from discussion here.

new topics are introduced in extra-clausal position, they become discourse-old in the remainder of the clause. Judging from this example, Givón's pronouncement that LD never introduces new topics appears too categorical. Lambrecht specifies that left-dislocated topics may be discourse-new, but are unlikely to be completely hearer-new (1994: 183):

> In both constructions [detachment or LD and presentational constructions], a referent is promoted from non-active to active status, and both constructions serve to establish a new topic. But while in the presentational sequence the referent of the NP is brand-new or at least unused, in the detachment case it is usually cognitively ACCESSIBLE. [...] The detachment or marked topic construction can then be defined pragmatically as a grammatical device used to promote a referent on the Topic Accessibility Scale from accessible to active status, from which point on it can be coded as a preferred topic expression, i.e. as an unaccented pronominal.

Prince's notion of "simplifying LD" may also be understood differently, not referring to the processing of old vs. new information, but in more general terms of processing syntactically complex constructions. Huddleston & Pullum provide the following example (2002: 1409):

(4.4.8) The people next door, the police have just arrested their son on a drugs charge.

Huddleston & Pullum propose that the main function of LD in this example is "simplifying" because "it serves to avoid having an awkward and complex genitive construction: ?*the people next door's son*" (2002: 1409). It is arguable whether this example should not better be analyzed as a subtype of the unlinked-topic construction (such as the double-subject construction "in which the topic and the subject stand in a possessive relation to each other", e.g. "*And **this guy**, **his fishing pole** fell down in the water*" (Lambrecht 2001b: 1059)).

Prince's second type is called "Poset LDs" because of their "set-inference triggering function" (1997: 126) which is similar to that of topicalization constructions:

> 'Poset' Left-Dislocation serves to trigger an inference on the part of the hearer that the entity represented by the initial NP stands in a salient partially-ordered set relation to some entity or entities already evoked in the discourse-model. (*ibid.*)

Recall from the chapter on topicalization that partially ordered sets or "posets" include such relations as part-whole, type-subtype, entity-attribute etc. Example (4.4.9) below illustrates this function of LD; speaker A, a student, tells B about her sister:

(4.4.9) A: Well you don't have a problem with that
artistic people don't
but people that have been in involved in sciences or <,> or a thing like that **they** just don't have creative minds
My sister for instance
She she just doesn't have any kind of a creative mind at all
And anything that's just not black and white facts you know scientific facts or mathematic facts she can't get her head round **it** at all
(ICE-GB: S1A-037 #036–41)

Speaker A finds several ways of expressing that her sister *doesn't have any kind of a creative mind at all*: if something is not *black and white,* such as *scientific facts or mathematic facts,* it is not accessible to her. These terms then clearly are in a poset relation, which licences the LD construction in this case.

To summarize: researchers who have dealt with left-dislocation are not unanimous about the information status that the left-dislocated topic may have, and are consequently equally divided on the pragmatic function(s) of the construction. Lambrecht's characterization of dislocation as a "marked topic construction" (1994: 183) probably captures the smallest common denominator of the preceding theoretical considerations. Advances in this area will most likely come about when researchers move away from decontextualized examples and take data from actual language use into account.

4.4.3 Right dislocation: Form and function

Again, right dislocation has a surface similarity to another construction that places a constituent in postverbal position, namely the existential construction, but the presence of a coreferential resumptive pronoun in the clause distinguishes RD from the existential. It is also intuitively obvious "that right-dislocation not only permits, but in fact requires, the dislocated NP to represent information that is given in some sense" (Birner & Ward 1998: 146). Pronouns by definition code referents that are accessible to the hearer in the discourse context, therefore the pronoun, which is coreferential to the dislocated NP, must designate a referent that has already been evoked in the discourse. In a sense, then, the right-dislocated NP is redundant: "this apparent inefficiency characterizes all instances of RD, since in all cases a referent is first coded in pronominal form before it is expressed in ANTITOP position" (Lambrecht 2001b: 1076).[55] One functional explanation that has been pursued for RD assumes that the construction represents some kind of "repair"

55. "ANTITOP" is Lambrecht's term for right-dislocation.

strategy, or "afterthought": the sentence-initial pronoun is felt to be in need of "clarification of reference" (Huddleston & Pullum 2002: 1411). The right-dislocated constituent is added by the speaker to disambiguate the reference of the pronominal form. However, "afterthought" RDs such as the one below are quite rare:

(4.4.10) I met **him**, your brother, I mean, two weeks ago.
(Lambrecht 2001b: 1076)

In this example, the afterthought *your bother, I mean* is indeed necessary to disambiguate the reference of the pronoun; unlike dislocated NPs in 'regular' RD constructions, the afterthought NP is accented. Lambrecht further points to examples like the following:

(4.4.11) They$_i$ take a long TIME, [them]$_i$ (In conversation about how to cook sweet potatoes) (2001b: 1076)

Examples such as this where the dislocated NP is realized by another pronoun clearly do not involve "clarification of reference". Lambrecht further notes (1994: 203, emphasis in the original):

> It is important to realize that the detached constituent in the antitopic construction does NOT express an AFTERTHOUGHT in the proper sense of this word, as has often been claimed. [...] A speaker who uses an antitopic construction is normally fully "aware" that the mere mention of the unmarked topic pronoun in the clause is insufficient for the hearer to understand who or what the proposition is about.

Two further examples from ICE-GB may serve to support the claim that RD is not primarily a "repair" strategy: in example (4.4.12), the discussion centres around the hazards of nuclear energy when A makes his rueful remark, and it is obvious that the dislocated NP *the human race* is neither intended nor needed to disambiguate the first person plural pronoun.

(4.4.12) **We** learn the hard way though don't **we** the human race
(ICE-GB: S1A-088#163:1:A)

In example (4.4.13), speaker C has even less need to clarify the reference of the resumptive pronoun *it* in his last sentence because the topic of this particular stretch of discourse, *the boat*, is well established and has been repeatedly evoked.

(4.4.13) The boat seats uh I mean sleeps fifteen hundred you see
And but they'd all gone on earlier it appears
and there it was completely empty [unclear-words]
so you've got to try and elbow your way through
There was nobody [unclear-word] on the boat <,>

> And in the restaurant there was all this space and the place and the thing was absolutely crammed
> There were hundreds of cabins
> 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#064–71:C)

Examples such as these show that unlike LD, right dislocation does not serve any topic-marking or topic-promoting function. Lambrecht points to "the similarity between topics and **vocatives**" (2001b: 1076) and notes that

> the ANTITOP vocative serves to secure the continued attention of an addressee, i.e. to maintain a given relation between a referent and a proposition. (*ibid.*)

Givón, on the other hand, suggests to analyze RD as a clue for other participants in the conversation to initiate a topic shift. He considers RD as a

> chain-final topic-marking device. [...] Given the strong indication of chain-final use of R-dislocation, it may be possible to view it as a **cataphoric** (rather than anaphoric) device, signaling the end of a thematic chain. It may thus alert the hearer to the likelihood of **topic switching** in the next clause. (1993: 215, emphasis in the original)

Givón's suggestion has some plausibility; if we return to example (4.4.13) above, we find that speaker C's turn which contains the right-dislocated NP is indeed discontinued. Another speaker, who has obviously shared C's experience on the boat, is first elaborating on C's account and then changing the overall topic:

(4.4.14) D: Even off the boat
You could walk to our cabins and not meet a soul
To complete the story Polly and I got up in the middle of the night
(ICE-GB: S1A-021#074–78)

Both explanations for RD might then be subsumed under the broad label "discourse management". We will see below in how far the discourse functions ascribed to left- and right dislocation play a role in the interpretation of the ICE-India data.

4.4.4 Dislocation: Corpus evidence

The ICE-GB grammar has a functional node DEFUNC (detached function), which "is applied to parenthetical clauses and vocatives" (ICECUP). Searching the direct conversation files for DEFUNC,NP turns up 849 hits; these include 32 examples of left dislocation and 50 examples of right dislocation.[56]

56. The categories DEFUNC,AJP (adjective phrase) and DEFUNC,CL (clause) did not contain any relevant examples.

The manual coding of relevant constructions in ICE-India was relatively straightforward. All examples retrieved were first assigned to either LD or RD. Within the LD category, I made a further distinction according to the syntactic function of the left dislocated constituent. The category of subject LD was again further specified by introducing a subcategory of "extended subject LD", i. e. syntactically complex left dislocated subjects such as subjects modified by a relative clause. This was done to test the hypothesis that LD serves a "simplifying" function as indicated above, helping the speaker to reduce the syntactic complexity of sentence structures by creating two separate processing units. Finally, I reconsidered my analysis for some ambiguous tokens with some surface similarity to dislocation constructions. The following is an example of an ambiguous construction which was eventually excluded from the dislocation category:

(4.4.15) A: Secondly you <.> shou </.> you should solve the numerical ability
B: Ah
A: Means <.> ma </.> <,> calculation should be finished off you know first only
B: Yeah
A: Because uh <,> after after <,> means reasoning after you try after reasoning <,> uh you'll be you'll get some difficulty in solving the <,> numerical ability <,> uhm <,> because **the calculation** sometimes will <,> uh **they** will be difficult (ICE-IND: S1A-071#81–85)

Speaker A's final turn does not represent a continuous utterance; rather, A starts off with *because **the calculation** sometimes will...* and then makes a fresh start after the pause and the hesitation marker: *uh **they** will be difficult*. The discontinuous syntax of the utterance then precludes the interpretation of the pronoun as resumptive.

We arrive, then, at Table 4.24 for the frequency and distribution according to syntactic contexts of dislocation constructions in both corpora.

Table 4.24. Frequency of LD and RD constructions in the ICE-GB and ICE-India direct conversation files (absolute figures and frequency per 10,000 words)

Dislocation construction	ICE-GB	ICE-India
left dislocation (total)	32 (1.56)	310 (14.18)
subject LD (incl. extended subjects)	28 (1.36)	276 (12.63)
extended subjects	10 (0.48)	27 (1.23)
object LD	4 (0.19)	28 (1.28)*
adverbial LD	–	6 (0.27)
right dislocation	50 (2,43)	14 (0.64)
total	82 (3,99)	324 (14.82)

* This figure includes prepositional objects and one example of a dislocated subject complement.

Again, the frequency of dislocation constructions in the two corpora is markedly different: taken together, left and right dislocation is four times as frequent in the ICE-India conversation files as compared to ICE-GB. Interestingly, it is the high number of LD tokens which accounts for this imbalance: LD constructions in ICE-India outrank those in ICE-GB by the factor nine, while ICE-GB features a higher ratio of right dislocation (3.8 to 1).

Biber et al. state about the distribution and frequency of dislocation across registers that both "types of dislocation occur over 200 times per million words in conversation and fictional dialogue, but very rarely in written prose" (1999: 957). Their figure amounts to two tokens per 10,000 words; the only other quantitative study that I am aware of provides a figure of 177 LD constructions in a corpus of 250,000 words, i.e. a frequency of 7.08 tokens per 10,000 words (Gregory & Michaelis 2001: 1678f.).[57] My figures for the ICE-GB conversations are then much lower in comparison, whereas the figures for LD (but not RD) in the Indian conversations are noticeably higher. The following sections will elaborate on this plain quantitative statement and consider the discourse contexts of both LD and RD constructions in more detail.

4.4.4.1 *Left dislocation in ICE-India and ICE-GB.* A closer look at dislocation constructions in the two corpora reveals some interesting differences in distribution. Turning first to left dislocation, we hardly find a difference concerning the dominance of left-dislocated subjects. 28 or 87.5% and 276 or 89.03% of all relevant examples in ICE-GB and ICE-India, respectively, have left-dislocated subjects, e.g.:

(4.4.16) Sir the reason is that <,> uh the father sometimes **he** is unemployed
(ICE-IND: S1A-074#175:1:B)

(4.4.17) B: No how did <,> how did you learn Hindi <,>
A: So **Hindi it** was compulsory for us <,> in uh
B: In the school
A: Yes yes in the school (ICE-IND: S1A-076#111–114)

The dominance of subject LD over object LD constructions is easy to explain. As noted above, an object can alternatively be marked as sentence topic by preposing, a syntactic option that is obviously not available for constituents which already occupy the sentence-initial position. Example (4.4.17) illustrates the by now

57. Gregory and Michaelis' study is based on the Switchboard Telephone Speech Corpus of American English. Biber et al.'s corpus data for the frequency of dislocation is based on BrE conversation: the noticeable difference in token frequency may reflect different usage patterns in BrE and AmE, but to investigate this further is beyond the present study.

well-established pattern of creating topic continuity by repeating the immediately preceding topic NP; I will come back to further examples of this discourse strategy below.

A striking difference in distribution is found in the proportion of "extended subject" LDs, i.e. subject NPs modified by a relative clause or some other clause, or consisting of several coordinated NPs. The ICE-GB data contain ten such left-dislocated extended subjects: eight tokens of a subject followed by a relative clause and two further tokens of a "heavy NP", e.g.

(4.4.18) The people we were staying with **they** <,,> uh cooked us a traditional Normandy dinner <,,> (ICE-GB: S1A-009#125:1:B)

(4.4.19) But uhm <,,> some people when describing people they don't like **they** just come out with this <,> kind of chain a bit like mine I suppose <,> chain of venom <,> (ICE-GB: S1A-037#239:1:B)

Ten extended subject LDs in ICE-GB amount to 31.25% of all occurrences of LD or 35.7% of all subject LDs; by contrast, extended subject LDs in ICE-India only contribute 8.79% to all LDs or 9.78% to the subject LDs, e.g.:

(4.4.20) The people who are <,> employed in uh <,,> very <.> luc </.> lucarious job <,> **they** will become <,> another upper class <,>[58]
And these people who are <,> employed in skill work **they** will become <,> the lower class <,> because <,> uh <,> the casteism <,> the origin of casteism was on that basis <,> in the beginning <,>
(ICE-IND: S1A-057#106–107:A)

(4.4.21) Then uh <,> there are so many serious <,> ragging <,> done by the seniors in north <,,>
But whereas in south like I feel the lecturers and the professors <,,> whoever in-charge then the principal **they** are <,,> controlling the students in that point of view <,,> (ICE-IND: S1A-090#115–116:B)

(4.4.22) And the students who come out with a degree MMS <,> I understand that there is a report that has been received <,> uh from different firms <,> that the students of BITS Pilani <,> specially MMS candidates <,> **they** are prepared to soil their hands <,>
(ICE-IND: S1A-023#185:1:A)[59]

Further, the ICE-India data feature a handful of examples for LD with a clausal subject, a construction that is completely absent from the ICE-GB data:

58. *Lucarious*: the speaker probably meant to say *lucrative*.

59. *BITS Pilani*: Birla Institute of Technology and Science in Pilani, an academic institution with a very high reputation; *MMS*: Master of management Studies.

(4.4.23) But uh <,> to go abroad uh **it** is very difficult [one or two words] because uh <,> the problem of language <,> the financial background everything comes in the way <,,> (ICE-IND: S1A-025#119:1:A)[60]

(4.4.24) Like <,> because this knowledge now <,> now a days what I'm learn learn <,> what we are learning **it** is twenty years old (ICE-IND: S1A-073#96:1:A)

This difference with regard to the comparative frequency of complex subject LD constructions is remarkable for two reasons both having to do with the "simplifying" function of LD as described above: first of all, if "simplifying" is understood to refer to the reduction of syntactic complexity, promoting ease of processing, then BrE speakers are much more likely than IndE speakers to resort to subject LD as a simplifying strategy. If it were indeed the case that New Englishes "favour a 'pragmatic' word order more than a strict 'syntactic' SVO order" (Mesthrie 2008: 632, cf. also 4.3.6 above), then this result is unexpected.

The second reason for highlighting the lower frequency of extended subject LD constructions in spoken IndE prefigures the discussion about possible motivations. It has been claimed for IndE that using an LD construction with dislocated subjects that are modified by a relative clause may be derived from substrate influence: Sridhar has noted that "Dravidian speakers of English typically avoid clausal relative clauses and instead resort to a variety of alternative structures, including [...] topic-comment structures" (1992: 144), e.g.:

(4.4.25) Anybody worshipping with devotion, they will be rewarded. (*ibid.*)

According to Sridhar, this example from an IndE speaker whose mother tongue is Telugu illustrates a common strategy to preserve the structural constraints on relative clause formation common to the Dravidian languages (1992: 144):

> The preferred mode of relative clause formation in Telugu, as in all Dravidian languages [...] is participialization. The relativized verb is turned into a participle, the coreferential subject deleted, and the structure is placed to the left of the head noun (in the main clause). This strategy satisfies three fundamental syntactic principles of Dravidian: (i) subordinate clauses precede the main clause, (ii) there is only one finite verb per sentence, and (iii) discontinuous constituents are to be avoided. Note that all these principles are violated (the first one frequently, and the other two necessarily) in a typical relative clause in English.

Example (4.4.25), even though the speaker uses a relative clause, adheres to principle (iii) of Dravidian syntax as described above: "the sentence is reformulated as a

60. The speaker is talking about unemployment in India and people who are leaving the country in search of jobs.

topic-comment structure, preserving the tendency to resist interrupting normal word order" (*ibid.*). Again, if substrate influence from the Dravidian languages were a major contributing factor to the high frequency of left dislocation in IndE, we would expect the percentage of left-dislocated extended subjects to be much higher, and especially much higher in comparison to BrE. Table 4.25 provides the figures for all instances of left dislocation in correlation to speakers' mother tongue. To be sure, these figures pertain not only to LD with extended subjects, that is, the specific syntactic context which has been identified by Sridhar as being particularly susceptible to substrate influence for Dravidian speakers of IndE, but to all LD constructions: the quantitative evidence is then only indirectly applicable. However, it is significant that IndE speakers of all the Dravidian languages taken together are underrepresented with respect to the occurrence of LD

Table 4.25. Frequency of LD constructions in the ICE-India direct conversation files according to speaker's mother tongue

mother tongue	count	percent	in the corpus (%)
Marathi	93	30.29	20.79
Kannada	66	21.50	19.92
Tamil	35	11.4	12.86
Punjabi	20	6.51	6.64
Malayalam	12	3.91	7.88
Oriya	10	3.26	2.07
Telugu	13	4.23	7.47
Hindi	12	3.91	5.81
Marwari	6	1.95	0.83
Bhojpuri	1	0.33	0.41
Bengali	7	2.28	3.73
Assamese	3	0.98	0.41
Naga	6	1.95	0.83
Kashmiri	5	1.63	1.25
Konkani	3	0.97	4.15
Angami	2	0.65	0.41
English	2	0.65	0.83
Manipuri	1	0.33	0.83
total	**307***	**100**	**99.99**
Indo-European		56.02	48.98
Dravidian		41.04	48.54
Other		2.93	2.48

* Information about speaker's mother tongue was only available for 307 tokens.

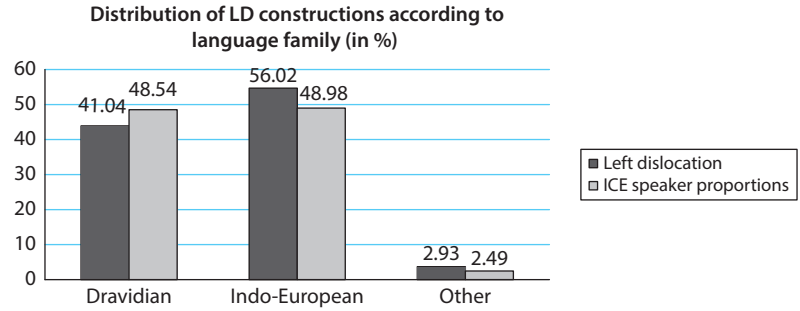

Figure 4.26. Distribution of LD constructions in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

constructions. As Table 4.25 shows, speakers of Kannada are slightly overrepresented, whereas speakers of Tamil are slightly underrepresented and speakers of Malayalam and Telugu are clearly underrepresented with respect to the use of dislocation. On a larger scale, then, Sridhar's observations about structural convergence within the realm of relative clause formation by a speaker of a Dravidian language cannot be extended to the IndE speech community as represented in ICE-India.

To return to our actual LD data: if we shift our emphasis from the purely syntactic contexts of LD constructions to the discourse contexts in which they occur, we encounter a pattern familiar from preceding chapters. A substantial number of LD tokens serve the function of creating topic continuity by repetition of the immediately preceding topic NP (cf. Table 4.27).

The examples (4.4.26) to (4.4.30) illustrate the prototypical discourse context for this function of LD: a speaker reacts to another speaker's questions by repeating the salient NP in an LD construction. Examples (4.4.26) to (4.4.28) show that this kind of discourse context triggers both subject and object LD.

Table 4.27. Frequency of LD constructions in the ICE-India direct conversation files with repetition of an immediately preceding topic NP (absolute figures and as percentage of the overall figure)

Dislocation constructions in ICE-India	total	NP repetition
left dislocation (total)	310	55 (17.74%)
subject LD	276	41 (14.86%)
object LD	28	11 (39.29%)
adverbial LD	6	3 (50%)

(4.4.26) D: How do you find the mess food over here Madhumita?
A: Mess food <,> **it's** a bit too hot I think so
(ICE-IND: S1A-007#9–10)

(4.4.27) B: Uh when did you <,> get <,> this job <,>
A: I got this job <,> in nineteen eighty-six
B: Nineteen nineteen eighty-six <,>
And uh when did you finish your <,> post graduate
A: My post graduation degree I finished **it** in <,,> uh mid June nineteen eighty-six (ICE-IND: S1A-009#60–64)

(4.4.28) 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#289–292)

In example (4.4.29), speaker B has just described her new golden earring in great detail when speaker A interrupts with *I thought you have bought some ordinary thing like* – not a question, but still a speech act which prompts speaker B to a reaction involving an LD construction.

(4.4.29) A: Ah <,> I thought you have bought some ordinary thing like
B: Ordinary thing I wouldn't have told you about **it** also <,> just because uh <,>
A: Okay <,> (ICE-IND: S1A-040#88–90)

Finally, example (4.4.30) shows that salient topics can also be expressed as clausal subjects.

(4.4.30) B: You are the only person staying in your room or uh <,>
A: Yeah I stay alone <,>
B: To be alone <,> **it's** more comfortable you know <,> staying alone how you [one word] (ICE-IND: S1A-016#174–176)

The observation that left dislocation is also employed to establish topic continuity is not as striking as it was in the context of the existential construction, which by definition is supposed to be the sole reserve of discourse-new referents. Recall from the discussion above that there is no agreement about the information status that a dislocated constituent may have. Scholars such as Prince include the introduction of discourse-new referents among the functions of LD (her "simplifying" function, cf. 4.2.3.2 above), whereas many others, Givón and Lambrecht among them, point out that the referent of the left dislocated constituent is typically already accessible in the communicative situation. However, the function of topic

repetition has so far not been described for LD constructions in other varieties of English.

There are only three examples where LD occurs not as a response to a question, but in the actual question. In example (4.4.31), the topic of school sports competitions has been mentioned previously and is now brought back as a topic.

(4.4.31) So inter-school sports when are **they** normally
(ICE-IND: S1A-003#230:1:A)

Similarly, in example (4.4.32) the LD construction establishes the unambiguous reference to the topic *Gardish* after another possible topic was introduced by another speaker.

(4.4.32) A: Uh which film he had put yesterday <,>
B: *Gardish* <,> okay nice film
C: Ours uh there was uh *Maine Pyar Kiya* <,>
A: In yours <,>
C: Ha <,>
A: *Gardish* how **it** is <,>[61]
B: Yeah nice movie <,> horrible movie <,,> you just watch it once <,>
(ICE-IND: S1A-051#284–290)

Examples such as these can easily be categorized in accordance with Gundel's suggestion that LD marks topics that are reintroduced into the discourse after a topic-shift.

Another set of examples for both subject and object LD can be labelled "poset LDs" in Prince's sense as discussed above. In example (4.4.33), the speakers talk about family matters and close relatives; the question *What about your sisters* triggers speaker B's enumeration of her three sisters' occupations, each of them introduced into the discourse by an LD construction.

(4.4.33) A: And uh what about your brothers <,>
What they're doing now <,>
B: My first brother is doing <,> uh B Com first year <,> and the other one is in eighth
A: What about your sisters <,,>
B: My first sister <,> uh **she's** doing beauty parlour course <,> and she has started already
And second sister uh **she** is uh doing <,> uh uh Engineering college <,> and third sister **she** is in B Com final year
(ICE-IND: S1A-053#31–36)

61. *Maine Pyar Kiya:* Bollywood love story released in 1989; *Gardish*: Bollywood crime movie released in 1993.

In the next example, the set relation between the established referent or entity and the referent(s) or entity/entities introduced by way of an LD construction is less transparent than in the first example. However, the *september issue* still belongs to the set of issues of the *computer magazine*.

(4.4.34) B: And what about this uh <,,> what is called as uh sending programmers to USA as a cheap labourer
A: Body surfing
B: Yeah body surfing I've read that in computer magazine
A: *Haan haan* <,>
B: September issue <,> they have send **it** <,>
(ICE-IND: S1A-045#117–121)

The speaker in example (4.4.35) talks about the difficulty of being admitted to university for a medical degree when most of the available places on a course are reserved for students from lower castes. Again, the whole – *hundred seats* – is contrasted with a part – *remaining seventy-six* – with the help of an LD construction.

(4.4.35) A: Out of hundred <,> uh seats <,> uh only twenty-four seats are open <,,>
Remaining seventy-six <,,> **they** are alloted to <,> this category or other category <,>
Only twenty-four seats are open to <,> these students <,,>
(ICE-IND: S1A-089#104–106)

A final interesting case of "poset LD" is example (4.4.36) below, featuring the only example of a negated dislocated constituent in the data.

(4.4.36) A: And uh students are allowed to write uh <,> uhm either in Kannada or English
B: Uhm
A: And even in science also that provision is there <,> But **none of the students they** write in Kannada <,> They're/> since they are taught in English <,> uhm <,> so they will answer in English
(ICE-IND: S1A-076#224–228)

Prince's category of "simplifying" LDs which serve to introduce a new referent into the discourse is also represented in ICE-India. The thematic structure in example (4.4.37) is quite interesting: the speaker first introduces a new topic, a person called *Smita* who seems to be known to both participants in the conversation, in a declarative sentence with its canonical old-before-new thematic structure. The shift to a new topic referent is then achieved with an LD construction: Smita's *friend Laxshmi* becomes the new topic which, once established, can then "be coded as a

preferred topic expression, i.e. as an unaccented pronominal", to use Lambrecht's expression once more (1994: 183). This example as well as the following one both support Lambrecht's point that the dislocated constituent is unlikely to be brand-new. *Laxshmi* as the topic referent in example (4.4.37) is already anchored in the discourse context by virtue of being *Smita's friend*, i.e. by virtue of the possessive construction which establishes a set relationship between *Smita* and *Laxshmi*.

(4.4.37) A: See yesterday I ran into Smita <,>
So uhm <,> her friend Laxshmi **she** was suffering from <,> her father was suffering from asthama <,> (ICE-IND: S1A-037#1–2)

In example (4.4.38), the LD construction *Poor Ashwini **she** didn't get admission in Delhi no* marks a rather abrupt topic shift; prior to speaker A's remark, the whole conversation has centred around the weather and *Ashwini* has never been mentioned before. Still, *Ashwini* must be known to the other participants in the conversation and therefore accessible as a topic, since they all enquire readily about her failure to gain admission to Delhi University in the remainder of the conversation.

(4.4.38) B: Today the weather is so pleasant *no* <,>
A: Pleasant <,>
A: Yeah
A: I don't know
B: It is very nice <,> just like a million
A: Come on <,> four four [one word] <,>
A: I am getting so [one word] especially <,>
C: You could
A: The fans are working
B: Yeah <,,>
A: Poor Ashwini **she** didn't get admission in Delhi *no* <,> because of the less percentage that she got she has got <,,>
(ICE-IND: S1A-070#313–323)

However, from a cursory look at the data from both ICE-India and ICE-GB, it seems that "simplifying" LDs are a minority option among the possible types of LD constructions proposed in the literature.

One final point worth mentioning is that not only subjects and objects, but also adverbials can be dislocated in spoken IndE, even though this seems to be a very marginal phenomenon. There are only six examples of adverbial LD constructions among all relevant examples in the ICE-India data and none in ICE-GB. The two examples quoted here feature the familiar pattern of topic repetition. Speakers A and B in example (4.4.39) are two young female students

who are continuously teasing each other throughout the conversation. In this excerpt, speaker A is ridiculing her friend for falling off her motorcycle:

(4.4.39) B: Yeah Yeah I fell down <,>
A: Where you have fallen down <,>
B: I fell down at my native place <,>
A: Near your native place <,> **there** also you have done same mistake uh <,> like uh riding in front of all guys like posing <,> you've done *na* <,> [laughter] (ICE-IND: S1A-049#7578)

In example (4.4.40), speaker A is trying to keep a conversation going with a rather monosyllabic speaker C, clinging to every bit of information that speaker A is ready to offer:

(4.4.40) A: Studying <,> what you are studying <,>
C: Commerce <,,>
A: In which college <,,>
C: I'm from Junior <,>
A: Junior college <,> how is staff **there** <,>
C: Fine <,,> (ICE-IND: S1A-051#15–20)

To summarize: relating our overview of the form and function of left dislocation in spoken IndE with the results for topicalization in the preceding chapter, we come across some interesting parallels. For both topicalization and left dislocation, the most obvious finding is that the two phenomena are much more frequent in IndE than in BrE. Topicalization in IndE shows a much wider range of innovative syntactic contexts than left dislocation in IndE, which can largely be analyzed along the explanatory parameters that have been proposed in the literature. Both constructions then share one specific usage which is absent from other varieties of English, namely that of providing an explicit link to the discourse context by repeating the immediately preceding salient NP. This property of LD will be discussed in more detail in 4.4.5 below.

4.4.4.2 *Right dislocation in ICE-India and ICE-GB.* With only 15 examples, right dislocation is a rare construction in ICE-India, both in comparison to the number of LD tokens in the same corpus and to the occurrence of RD in ICE-GB, where RD constructions actually outnumber LD constructions. Five of the 15 RD tokens occur in questions – interestingly, there is no single parallel example from ICE-GB for RD in interrogatives. Both example (4.4.41) and example (4.4.42) illustrate that right dislocation does not serve the function of "clarification of reference", as Huddleston & Pullum (2002: 1411) suggested. Speaker A in example (4.4.41) enquires about the location of a particular shop that has just been mentioned.:

(4.4.41) B: Such as I [laughs] [one word] Parmar's shop as well as <,>
A: Ah where is **it** Parmar's <,,>? (ICE-IND: S1A-075#29–30)

In example (4.4.42), the participants in the conversation are discussing their neighbours, who are both called *Patil*, and again the name *Patil* has just been mentioned so that there is no scope for ambiguity of reference.

(4.4.42) Ahn they <,> I thought that another one Patil is there a horrible he is I thought **that Patil** (ICE-IND: S1A-053#323:1:A)

One interesting observation concerns the realization of RD in spoken BrE. Ten out of the 50 examples for right dislocation in the ICE-GB data involve a right dislocated demonstrative pronoun rather than an NP; incidentally, all relevant examples feature non-referential *it* as subject, e.g.:

(4.4.43) It's a cultural thing **that** you know (ICE-GB: S1A-043#313:1:B)
(4.4.44) you know it's a wankers' table **that** <,> (ICE-GB: S1A-052#078:2:A)
(4.4.45) It all fits in this section **this** (ICE-GB: S1A-077#105:1:A)

This construction has so far not received much attention in the literature; Biber et al., who use the term "noun phrase tag" for right dislocation, state that the "noun phrase tag in such examples serves to emphasize the proposition, in much the same way as declarative tags" (1999: 958). Parallel examples are absent from ICE-India, even though there are three examples which feature the preferred syntactic context for right-dislocated pronouns, namely clauses introduced by non-referential *it*:

(4.4.46) The coconut plants are <,> actually **it** was our property **that coconut** <,> now it has become on all India (ICE-IND: S1A-035#236:1:B)[62]

It seems reasonable to assume that right dislocated demonstrative pronouns are functionally equivalent to tag questions, specifically to so-called 'attitudinal' tag questions which emphasize the speaker's proposition and do not elicit a response from the listener (cf. Chapter 4.6.2). We will see in Section 4.6.2.3 that spoken IndE generally favours indigenous tags over English-derived tags, which might explain the absence of noun phrase tags in the data.

4.4.5 Explanatory parameters

We have already seen in Section 4.4.4.1 above that left dislocation cannot be explained as a strategy of pattern replication typical for speakers of Dravidian languages. None of the other speaker variables has a statistically significant impact on

62. Context: the speaker remarks that coconut trees originally come from his home state Kerala, but have now spread all over India.

the distribution of LD constructions.[63] Some patterns are already familiar from preceding chapters, for example the overrepresentation of women and younger speakers with respect to a particular construction.

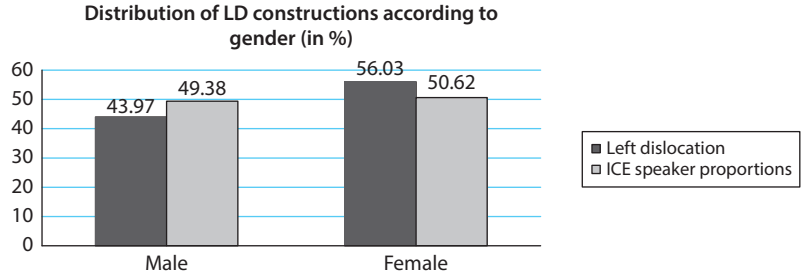

Figure 4.28. Distribution of LD constructions in the ICE-India conversation files according to speaker's gender

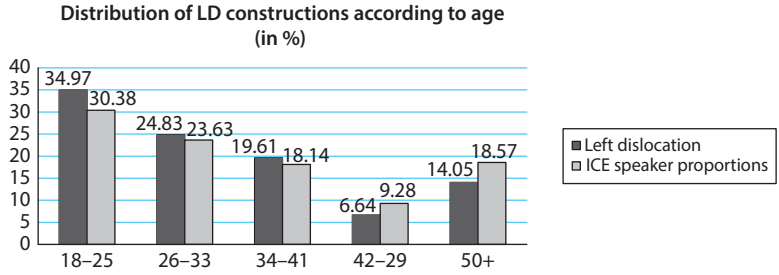

Figure 4.29. Distribution of LD constructions in the ICE-India conversation files according to speaker's age

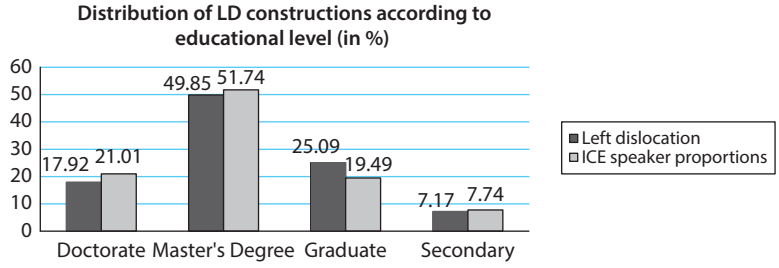

Figure 4.30. Distribution of LD constructions in the ICE-India conversation files according to speaker's educational level

63. P = 0.0664 for the variable gender; p = 0.0696 for age; p = 0.0839 for educational level.

More promising explanations are possible if we take the specific IndE discourse functions of both left dislocation and topicalization into account. The previous chapter has shown that topicalization in spoken IndE is almost ten times as frequent as in spoken BrE, and the figures for left dislocation in both varieties are parallel. Both constructions further overlap in a discourse function that is specific to IndE, namely creating cohesion in the ongoing conversation by repeating what the previous speaker has said. This strategy has been mentioned in Chapter 4.2.6.1 above as "elliptical repetition". Subbarao et al. explain the use of this discourse strategy for Hindi, but stress that it is a pan-Indian device to indicate politeness (1991: 46f.):

> A very common strategy to achieve positive politeness [...] is to repeat a part of what the other interlocutor has said. For example, consider the following piece of Hindi conversation:
>
> A: varma: sa:hib kya: ba:za:r gae hãĩ?
> Mr. Verma honorific suffix yes/no Q-marker market has gone
> 'Has Mr. Verma gone to the market?'
>
> B: ji: nahi:, ghar par hãĩ
> honorific marker no home at is
> 'No, sir, he is at home.'
>
> A: accha:, ghar par hãĩ
> I see home at is
> 'I see, he is at home.'
>
> The repetition of *ghar par hãĩ* [...] does not communicate any new information and should be considered as phatic communion. Yet it appeases the positive face of B and is a polite way of indicating that A has taken notice of what B said.

Given the fact that the discourse functions of both topicalization and LD in spoken IndE are similar to a considerable extent, we might question Birner & Ward's rather categorical distinction between the two constructions as not only formally but also functionally quite different (1998: 94). On one level, left dislocation in spoken IndE can simply be seen as a topic-marking device for subjects, i.e. constituents that already occupy the sentence-initial position and are thus not amenable to preposing.

It is then again the South Asian "grammar of culture" (cf. Chapter 3.2.4) with specific syntactic strategies for expressing politeness that is the best candidate for substrate influence. On a purely syntactic level, substrate influence is ruled out: Indian languages generally favour zero anaphora for topics, so it is highly unlikely that the addition of resumptive pronouns is motivated by substrate influence. The examples of fronted topics in Indian languages discussed in Section 3.2.3 above

provide further illustration. None of the examples features a resumptive pronoun where the English translation has one, as can be seen in the Malayalam example (3.7), repeated here as (4.4.47):

(4.4.47) ravi varmayuṭe citram oru ameerikkakkaraan vaaŋŋi
Ravi Varma-GEN painting an American buy-PAST
'Ravi Varma's painting, an American bought it.'

Note that the Malayalam topicalization construction becomes an LD construction in the English translation, which is again evidence for the assumption that both constructions might not be as distinct as Birner & Ward would have it.

Still, if we only look for corresponding patterns on the level of syntax, we might miss the crucial fact that syntax is just one option to code information structure, and that languages may employ different linguistic devices to different degrees. Recall from the beginning of this chapter that Platt et al. (1984: 120) referred to subject LD as

> a perfectly legitimate device for emphasizing the subject, which is quite commonly used in educated varieties of other languages, e.g. in French
>
> *Ma mere, elle* est très élégante
>
> It is a particularly useful device for speakers of the New Englishes who do not make the same use of intonation for emphasizing as do some of the speakers of British English.

This statement, which appears so alluringly commonsensical, actually raises a host of problems: problems of definition and problems of explanatory adequacy. If, to begin with, we equate "emphasis" in relation to "intonation" with "prosodic prominence", then the relevant literature tells us that it is generally the focus of a sentence which receives prosodic prominence, and not the topic. In Lambrecht's terms (1994: 337):

> The overriding function of sentence accents is to indicate denotata with pragmatically non-recoverable relations to propositions, whether topical or focal.

To put it differently: if a constituent of a sentence is marked by prosodic prominence, it is either the focus of the sentence or a contrastive topic – new information on both accounts. The correlation between intonation/emphasis in the "Old" Englishes and dislocation in the New Englishes is then much less straightforward, and should be studied by taking the whole domain of topic and focus constructions in a particular variety into account. We will have ample opportunity to consider the different focus marking strategies that are available in languages in the two subsequent chapters.

4.5 Cleft constructions

Like the dislocation constructions which were the subject of the preceding chapter, cleft constructions are not included in Ward & Birner's (2004) overview of information structure and non-canonical word order because the construction involves much more than a re-arrangement of the linear order of constituents. Clefts "express a single proposition via biclausal syntax" (Lambrecht 2001a: 466), as in the following examples:

(4.5.1) When I was at home <,> more often than not **it was my mother** that cooked or sometimes my father <ICE-GB: S1A-059#169:1:B>

(4.5.2) **It's his Mum** falls in love with him (ICE-GB: S1A-006#128:1:A)

(4.5.3) **It wasn't him it was me** who was at fault <,,> (ICE-GB: S1A-050#084:1:B)

(4.5.4) Yeah **was it you** that told me that <ICE-GB: S1A-099#271:2:A>

Examples such as these are also known as *it*-clefts, and it is this subcategory of cleft constructions that has received most attention in the literature, both because of their form and their discourse functions. Another type of cleft is known as either *wh*-cleft or pseudo-cleft, e.g.:

(4.5.5) E: And that has also been <,> uh pointed out by many <,> and not every teacher can be really made objective
F: No **what I really wanted is** to make the objective elements of the teacher comes into play then <,>
But **what I still feel is** <,> you've different means and methods by which you can evaluate the students to continue the assessment <,>
(ICE-IND: S1A-011#142–144)

The term "pseudo-cleft" already indicates that the construction, despite some surface similarity with the *it*-cleft construction and regardless of its actual frequency, is somehow marginal to the category. In Huddleston & Pullum's terms (2002: 1423):

> The *it*-cleft can almost invariably be matched with an equivalent non-cleft, and it clearly represents a distinct construction by virtue of the special use of *it* and the special type of relative clause. The pseudo-cleft, by contrast, can very often not be matched with an equivalent non-cleft, and rather than constituting a distinct construction it simply represents the particular case of the specifying *be* construction where the element defining the variable is realised by a fused relative.

In the following, I will therefore concentrate on *it*-clefts. The reason for including a chapter on *it*-clefts in my study is admittedly not intuitively obvious: the only

varieties of English which have been noted so far as displaying an innovative us of cleft constructions are the Celtic Englishes (cf. Odlin 1997: 36–41, Filppula 2006: 526).[64]

For clefts in Hiberno-English and Celtic Englishes generally, the higher frequency as well as the greater range of constituents that can be clefted is commonly attributed to substrate influence. However, nothing of the kind can be observed for IndE, and as we will see in the following, it is rather the absence of *it*-clefts in the ICE-India data that requires explanation. This chapter will therefore necessarily be much shorter than the others. Its main purpose is to complete the overview of syntactic means of coding information structure and then to lead over to other available means of focus marking.

4.5.1 Definition

Examples (4.5.1) – (4.5.4) above illustrate the general form that *it*-clefts take:

> *It*-cleft clauses have *it* as the subject of the matrix *be* clause, with the relative clause appearing in extranuclear position at the end. The *it* in subject function can be thought of as a place-holder for the variable, which is defined in a relative clause that is not syntactically part of the subject. (Huddleston & Pullum 2002: 1416)

The examples further illustrate the range of variation found with *it*-clefts. In examples (4.5.1) and (4.5.4), the relative pronoun introducing the relative clause is *that*, even though the antecedent is animate. In spoken English, it is also quite common to omit the relative pronoun altogether, as example (4.5.2) shows. Another common bone of contention for prescriptivists is exemplified in example (4.5.3), namely "the pronoun problem", i.e. the case of the pronoun in copular clauses such as *It's I/me* and, by extension, in it-clefts.[65] Whereas prescriptive grammars insist on the nominative or subjective case for the pronoun, modern grammars acknowledge that the choice of pronoun is largely a matter of style rather than correctness. Huddleston & Pullum, for example, point out that *it*-clefts and copular sentences are among the "constructions where nominative and accusative are in alternation [...] the nominative is restricted to formal (or very formal) style, with the accusative appearing elsewhere" (2002: 459).

Cleft constructions belong to the repertoire of focus constructions, more specifically, they represent one option of realizing contrastive focus. In Drubig & Schaffar's terms (2001: 1079):

---

64. Mesthrie notes that predicate clefting (e.g. "*it's tiredness that tires me*") is common in West African Pidgin Englishes, but not found anywhere else in Africa or Asia (2008: 634).

65. For a more detailed discussion of the relation between copular clauses and the so-called 'truncated *it*-clefts' cf. Lange & Schaefer (2008).

> There is a growing consensus among students of information structure that two types of focus must be distinguished in terms of form and interpretation: presentational focus (or information focus, focus of assertion, rheme, usually wide focus) refers to a constituent which must be interpreted as new, or context-incrementing information, whereas contrastive focus (or identificational, operator focus, usually narrow focus) denotes a constituent that identifies a subset within a set of contextually given alternatives.

The notion of "contrastive focus" will become clearer when we compare a canonical sentence with its topicalized and its clefted version, respectively.[66] Sentence (a) displays the unmarked topic-comment structure, where the object CHAMPAGNE by default is the focus and is marked by prosodic prominence. In (b), the focus is preposed in a topicalization construction; the felicity conditions for this construction have been extensively discussed in Chapter 4.2.2. In (c), the focus is given further syntactic prominence by appearing as the focus phrase CHAMPAGNE in the matrix clause *It's* CHAMPAGNE of a cleft construction.[67]

a. I like CHAMPAGNE
b. CHAMPAGNE I like
c. It is CHAMPAGNE (that) I like

What distinguishes the meaning of sentences (a) and (c), then, is the additional presupposition in (c) that the speaker likes nothing else **but** champagne (at least given the current context). This particular function of clefts can be illustrated with another example from ICE-GB which appears distinctly odd at first sight:

(4.5.6) A: Bloody ***auk*** <,,>
B: There's nothing wrong with my ***auk***
It was the ***gauge*** that was the killer in the first place <,,>
(ICE-GB: S1A-010#124–126)

The two speakers are engaged in a game of Scrabble and are considering their options. When speaker B points out that *It was the **gauge** that was the killer in the first place*, she uses a cleft construction to single out one of the Scrabble words among all others on the board in front of her, the one word that is, among all other possible options (for example the ***auk***), solely responsible for the current state of the game in her opinion. Lambrecht's term "exhaustive specificational clefts" (2001a: 497) captures the pragmatics of the construction quite well.

66. Examples adapted from Lambrecht (2001a: 467).

67. The terminology follows Lambrecht (2001a).

4.5.2 Clefts in ICE-GB

It-clefts in the ICE-GB conversation files can conveniently be retrieved by searching for the word class label ITCLEFT, turning up 46 examples, which amount to a frequency of 2.23 tokens per 10,000 words. Nelson (1997) found overall 387 cleft construction in the whole ICE-GB, and only 40 tokens in the direct conversation files after eliminating some unclear cases. His statistics thus do not support the claim made by Quirk et al. (1985: 1385) that clefts are a very common feature of spoken English. Nelson found that "cleft constructions are typical of public discourse, and of public monologue in particular" (1997: 344), but adds that "typical" should be taken with a grain of salt, since the overall frequency of clefts is so low to begin with.

4.5.3 Clefts in ICE-India

As I have already indicated, clefts in spoken IndE are an even more marginal phenomenon than in ICE-GB: there are only seven examples in five texts (0.32 tokens per 10,000 words), and these seven examples only represent five different tokens: in example (4.5.8), the utterance ***it's they*** *who collect the fund* is first repeated once by the same speaker and then echoed by the other participant in the conversation.[68] The syntactic context of example (4.5.10) is further not quite clear from the context of the conversation, which finally leaves us with no more than four different unambiguous tokens (frequency 0.18 per 10,000 words). If we pit all seven tokens against the number of clefts in ICE-GB, then clefts in the BrE data are seven times more frequent than in the Indian data. If we only take the four unambiguous tokens in ICE-India into account for the comparison, then clefts in BrE are 12.4 times as frequent in comparison to IndE. The following is an exhaustive list of all examples:

(4.5.7) C: No but India is mostly a <,> patriarchal society
So **it's men** who play a more important role right <,> in all your decision making and <,> what about this <,> however modern we may get to but still the same <,> (ICE-IND: S1A-011#23–24:1:C)

(4.5.8) A: Of course the these cultural festivals are organised by students themselves
B: By students themselves <,>
A: Yes **it's they** who collect the fund <,> <{> <[> **it's they** who collect fund <,,> organise
B: <[> **It's they** who collect fund </[> </{>
(ICE-IND: S1A-023#179–182)

68. The textual markup in example (4.5.8) is as follows: overlapping strings are individually enclosed in square brackets: <[> ... </[>; the overlapping strings together are enclosed in <{> ...</{>.

(4.5.9) A: But basically I am a mathematics teacher but <,> uh in my school I teach uh English
C: Oh ho ho **it's fantastic job** you are doing *no* <„>
(ICE-IND: S1A-025#21–22)

(4.5.10) B: **It is some achievement** who has done <,>
(ICE-IND: S1A-045#46:1:B)[69]

(4.5.11) A: **It's a good time-pass** we have (ICE-IND: S1A-061#215:1:A)[70]

As said above, even given the overall rarity of it-clefts in ICE-GB, the extremely limited number of clefts in the corresponding ICE-India files is in need of explanation. Any further analysis of the form and functions of clefts in spoken IndE is precluded by the lack of examples.

4.5.4 Explanatory parameters

The observation that the occurrence of clefts in a language correlates with fixed vs. free word order has already been made by Jespersen (1937/1969) and elaborated by Lambrecht (2001a). Cleft constructions are clearly associated with languages such as English with a rigid word order. Lambrecht's principle captures "the formal motivation for the occurrence of clefts" (2001a: 488):

> The occurrence of cleft constructions in a language correlates with the degree of positional freedom of prosodic accents and syntactic constituents in that language.

To put it differently, clefts as a syntactic means of marking contrastive focus will occur in languages where the other two options for indicating focus are more restricted, namely "prosodic shifts (changes in the unmarked position of focus accents)" and "syntactic shifts (changes in the unmarked position of focus constituents)" (*ibid.*). Languages with a less rigid word order than BrE such as German may easily resort to a change in the canonical word order to place the focus constituent in a cognitively salient position; consequently, *it*-clefts are much rarer in German than in English. French, on the other hand, makes extensive use of clefts construction, and this is according to Lambrecht due to the combined impact of comparatively fixed word order and quite fixed rhythmic structure (1994: 224f.):

---

69. The syntactic context does not become clear from the context of the conversation.

70. Context: the speakers are talking about attending temples and other ways of passing the time.

> I believe that differences in the rhythmic structure of languages account at least in part for the use of particular focus-marking systems. For example it seems likely that in French the prevalent use of cleft constructions for the marking of focus differences is at least in part due to the fact that this language has both a relatively rigid constituent order and a relatively rigid rhythmic structure.

We have seen in the preceding chapters that word order in spoken IndE displays much greater freedom and can easily be accommodated to the expression of information structure. The role of the IndE "rhythmic structure" as a determinant of the type of focus marking – syntactic, prosodic, or morphological – will also play a role in the next chapter.

4.6 Utterance modifiers

The preceding chapters have dealt with various syntactic strategies that can be conveniently subsumed under the label "information structure"; such strategies belong to all intents and purposes to syntax proper.[71] The two topics treated in this chapter are generally assigned to the realm of pragmatics rather than syntax. Both focus markers and discourse markers by definition do not contribute to the truth value of a sentence. However, such a negative definition does not in itself justify treating both phenomena in the same chapter, let alone their inclusion in a study that has so far been concerned with information structure and non-canonical word order. One reason for including focus markers and discourse markers has already been mentioned in Chapter 4.1.3: both are devices available for discourse management, broadly conceived. There is a further reason for allocating an investigation of focus markers and discourse markers to a single chapter under the heading of "utterance modifiers". As already mentioned briefly in Chapter 2.3, the term was suggested by Matras (1998) for a class of items that are crosslinguistically particularly prone to borrowing in language contact situations. The class consists of

- adversative (and partly other) coordinating conjunctions;
- sentence particles, such as 'well', 'so', 'anyway';
- fillers, tags, and interjections;
- focus particles [...], including phasal adverbs, such as 'still', 'yet', 'already'. (Matras 1998: 294)

Matras argues that they have a common denominator in that they belong to "the grammar of directing" (1998: 325):

71. Cf. Lambrecht (1994: 1–13) for an elaboration of this point.

> these "utterance modifiers" participate in a regularized, closed set of structures on which speakers draw in order to direct hearer-sided processing of the propositional content of utterances and ensure the acceptance of propositional and interactional coherence in discourse. With interactional coherence I mean the harmonious continuation of negotiated speaker-hearer roles in a particular position in the discourse. Utterance modifiers thus contribute to a component of grammar that the speaker uses to DIRECT the hearer's reactions. (Matras 1998: 294f.)

IndE also features discourse markers familiar from other varieties of English, such as *well*, *actually*, *like*, *I mean*, and others. A brief look at raw frequencies already indicates that a closer comparative study of discourse markers would prove extremely interesting: *well* in all its functions occurs 364 times in the ICE-India conversation files and 1,729 times in the parallel ICE-GB data. If we eliminated all instances of adverbial *well* (including *as well*) from the data, we would probably still be left with at least ten times as many instances of *well* as discourse marker in BrE than in IndE, if not more. Valentine (1991) is the only study of an array of IndE discourse markers that I am aware of; she notes:

> Hindi words such as *acchaa* 'good', complementizer *-ki*, *matlab* 'meaning', *bilkul* 'absolutely', *suno* 'listen', *Thiik hai* 'okay', among others, are used spontaneously and naturally in the Indian speakers' English conversations. The use of these lexical borrowings demonstrates how the Hindi items have been assimilated into the English linguistic system. These bilingual speakers, especially the younger generation, use these terms by themselves, not always as a unit in combination with other Hindi items. (1991: 331)

However, of the discourse markers on Valentine's list, two do not occur at all in the ICE-India subcorpus (*bilkul* and *suno*), three further forms are negligible (*matlab*: two times; *theek hai*: five times; *-ki*: 15 times), and only *accha* with 202 tokens is reasonably well represented. *Matlab* seems to have been calqued as *means* rather than *meaning*, as the following examples show.[72] Speaker C in example (4.6.1) talks about the buses in Hyderabad, where the front of the bus is reserved for men and the back for women. Speaker A in example (4.6.2) inquires about the customers in her friend's new beauty parlour:

(4.6.1) Now they have <,> uh introduced before it wasn't like that <,> and still there is a problem people <,> **means** they come and sit <,> on the ladies seat and they won't they get up <,>
We have a lot of quarrel (ICE-IND: S1A-021#219–220)

72. Interestingly, *meaning* occurs only 15 times in the ICE-India conversation files, and then only as a noun.

(4.6.2) So most of your uh <,,> **means** <,> clients are <,> uh young girls or <,> middle aged ladies (ICE-IND: S1A-043#52)

The distribution and function of *means* would be worth a closer scrutiny, but is outside the scope of the present study. Chapter 4.6.2 will therefore be exclusively concerned with tags, more specifically the invariant tags *isn't it* and *no/na* as the most frequent and the most conspicuous discourse markers in IndE. Chapter 4.6.1 is devoted to the focus markers *only* and *itself*, which have developed innovative functions in IndE.

4.6.1 Focus markers: *Only* and *itself*

In his overview of morphological and syntactic variation in L2 varieties of English in Africa and South and Southeast Asia, Mesthrie notes: "In addition to occasional standard usage, *only* and *too* are focus markers in IndE and InSAfE [Indian South African English]" (Mesthrie 2008b: 632). We are dealing, then, with a phenomenon that is the exclusive property of New Indian Englishes. Mesthrie provides an example for the usage in question:

(4.6.3) I stay next door to my mother *only*. (= 'I live right next door to my mother') (1992: 55)

He adds that *only* "may also be used in its English English sense, limiting the meaning of a lexical item", but points out that in the example given above, "this limiting meaning is only a secondary nuance" (*ibid*.). However, simply labelling *only* and *too* 'focus markers' does not capture the innovative function they have developed in IndE, since they also mark focus in their "English English sense". We have already seen in the preceding chapter on cleft constructions that two types of focus are commonly distinguished, and this distinction is particularly relevant for an analysis of IndE *only*. For reasons that will be spelled out below, I will restrict my analysis to *only* and *itself*, disregarding *too* and *also*.[73]

4.6.1.1 *Definition.* Both *only* and *too/also* belong to the set of expressions that have been classified as "focus particles" (König 1991, 1993) or "focusing adverbs" (Nevalainen 1991).[74] König's general definition of focus particles runs as follows (1993: 979):

73. A recent comprehensive account of *also* in IndE is Fuchs (2012).

74. The label "circumstance adverbial" chosen by Biber et al. (1999: 780f.) disregards the specific function related to the focus structure of a proposition.

> Focus particles, and in fact all types of focusing, relate the denotation of a focus to a set of denotations of the same type. [...] In addition to establishing such a relation to alternative values, focus particles also typically either include or exclude such alternatives as possible values for the propositional schema in their scope.

Example (4.6.4) illustrates the effect of focus particles on the "focus-background structure of a sentence" (1993: 978). The speakers are organizing their "hostel day" and are discussing the preparations for the entertainment programme:

(4.6.4) B: Uh see we are having two dances okay
A: Uhm what type of dance <,>
B: And
Film hits **only** <,> (ICE-IND: S1A-049#13–16)

Speaker B's utterance *Film hits **only*** is a response to speaker A's question, enquiring which types of dance should be performed. *Film hits* in this utterance is clearly the focus, defined as "that string of expressions which is set off from the rest of the sentence by prosodic prominence and which is specifically affected semantically by the particle" (1993: 979). In this case, the identification of the focus is trivial, as there is no other potential focus around. However, given the positional variability of focus particles, the identification of the focus of a more complex sentence might be less than straightforward (cf. König 1993: 981). The addition of *only* then marks *Film hits* as the focus and relates the denotation of the focus to a set of alternatives, in this case other possible types of dance. Further, *only* explicitly excludes these alternatives from consideration: no other dances will be performed, *only* those that have appeared in Bollywood movies. In König's terms (1991: 98):

> A sentence with *only* presupposes the relevant sentence without particle and entails that none of the alternatives under consideration satisfies the open sentence obtained by substituting a variable for the focus expression.

When considering the contribution they make to the meaning of a sentence, focus particles fall into two broad categories. Expressions such as *only* then belong to the subgroup of exclusive or restrictive focus particles. Inclusive or additive focus particles, on the other hand, "include some alternative(s) as possible value(s) for the variable in their scope" (1991: 33). In example (4.6.5), speaker A from Orissa compares speaker B's home state Maharashtra favourably with her own:

(4.6.5) A: You've really I'm sorry in a very good way you've convinced me <,> and uh <,> I found the climate of <,> Maharashtra is very good people are nice soft spoken well behaved and uh <,> exactly like the Oriya people <,>
The dishes are **also** <,> uh quite similar to the <,> dishes of Orissa <,> and <,> the system you know the culture <,> the culture differs a little bit I think
B: Yeah (ICE-IND: S1A-004#83–85)

When speaker A compares the Orissa *cuisine* with Maharashtrian *cuisine*, stating that *The dishes are **also** <,> uh quite similar*, the focus marker *also* gives rise to the implicature that there might be other things that are *quite similar.* Such alternatives to the focus *the dishes* have already been mentioned in speaker A's turn, namely *the climate* and the *people*.

Unlike the additive/inclusive focus markers, restrictive/exclusive focus markers such as *only* have a syntactic correlate in the cleft construction. Both exclusive focus markers and the *it*-cleft mark contrastive focus rather than presentational focus (cf. Chapter 4.5). Kiss, who prefers the terms "identificational focus" for exclusive or contrastive focus and "information focus" for presentational focus, again states the difference between the two types of focus (2001: 1449f.):

> An identificational focus identifies a proper subset of a set of contextually or situationally given elements as such for which the action or state described in the sentence exclusively holds. [...] Whereas identificational focus expresses exhaustive identification, information focus merely marks the non-presupposed nature of the information it carries.

The typology of contrastive focus-marking within the realm of crosslinguistically attested realizations of focus constructions may be visualized as in Figure 4.31. In natural languages, focus may be realized by prosodic, morphological, or syntactic means, or a combination of these (cf. Drubig & Schaffar 2001: 1095–1100). Put differently,

> syntax is not the only formal level at which information structure is coded. What syntax does not code, prosody does, and what is not coded by prosody may be expressed by morphology or the lexicon. (Lambrecht 1994: 31)

Across languages, "prosodic prominence is perhaps the most frequently encountered focussing device" (Drubig & Schaffar 2001: 1080), an observation supported by Lambrecht, who seeks an explanation for this attested preference for prosodic focus marking in the "iconic relationship between pitch prominence and the degree of communicative importance assigned to the focal portion of a proposition" (1994: 224).

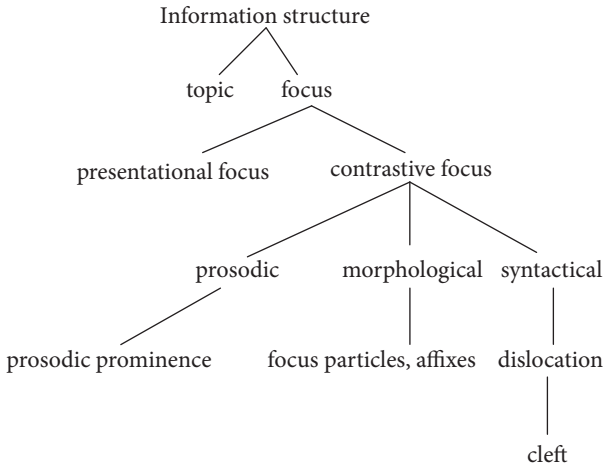

Figure 4.31. The domain of contrastive focus marking

The label "morphological" in Figure 4.31 should be understood as including both inflectional morphology and free lexical forms such as the English focus particles. We will see below whether the extreme rarity of *it*-clefts in spoken IndE noted in the relevant chapter is in some way compensated by a corresponding high frequency of lexical focus constructions involving the exclusive focus marker *only*. When Mesthrie claims that *only* is a focus marker in IndE and South African IndE, he does not have the exclusive sense in mind which is common in all varieties of English and which is a realization of contrastive focus. His example quoted above illustrates that *only* in the New Indian Englishes can also mark presentational focus. He further notes that *only* in this innovative usage is restricted to basilectal SAIE (1992: 56). His data also reveal a significant correlation between speakers' background languages and the occurrence of presentational *only*: only speakers with a Dravidian language background (Tamil or Telugu) had presentational *only* in their repertoire, whereas speakers of Indo-Aryan languages (Bhojpuri or Urdu) did not use *only* with the innovative meaning (1992: 157f.). We will see below whether the data for IndE show a similar pattern.

The fact that *only* does double duty as contrastive and as presentational focus marker in IndE is well-established (cf. Bhatt 2000, 2008). In Lange (2007), I showed that the intensifier *itself* has also developed a new function as presentational focus marker. In example (4.6.6), two college students are talking about their studies in general and their hostel facilities in particular. Speaker B is a bit envious of speaker A, who lives in a hostel with a gym, and encourages her to take full advantage of the facilities. When intensifying *itself* occurs following its focus, then the focus has

to be an NP in all other varieties of English;[75] the fact that focus *itself* in this case interacts with an adverbial already indicates that the new meaning of *itself* has also affected the combinatorial possibilities of the form.

(4.6.6) B: I think you should start going to the gym from now **itself** (ICE-IND: S1A-061#86:1)

Example (4.6.7), though unusual, may be taken to indicate that *itself* as presentational focus marker in IndE is already on its way to being grammaticalized. The lack of agreement between the focus *man* and the following focus particle indicates that *itself* is used as a focus particle like *only*, i.e. an invariant form.

(4.6.7) B: Okay okay <,> Indian culture says that women should work <,> yeah <,> isn't it? She should cook she should look after the children it's alright <,> But even if they women is are working as much as a man <,> she is earning the same <,,> uh monthly saving <,> as a man **itself** <,> (ICE-IND: S1A-087#165–167)

In the following, I will consider the corpus evidence for both uses of *only* and *itself* in more detail.

4.6.1.2 *Focus markers* only *and* itself *in ICE-India: Corpus evidence.* In the absence of contextual information, it is frequently difficult to classify all instances of *only* as either restrictive focus particle or presentational focus marker. One reason for this difficulty is the marked preference of IndE speakers for placing *only* in final position (following the focus) where most other varieties prefer the order focus particle – focus, e.g.:

(4.6.8) There are two things **only** <,> either you join those people <,> or you fight those people <,> (ICE-IND: S1A-083#167:1:A)

Here the context makes it perfectly clear that *only* follows its focus and has a restrictive sense. Contexts which involve some kind of quantification as the one above are generally very likely to involve the restrictive focus sense of *only*; this default assumption, however, proves to be too superficial, as the following example illustrates:

(4.6.9) A: Yeah <,> of course we <,> we are being charged also that much **only** for it *no* <,,> We are having more uh <,> hostel fees uhm <,> as compared to others colleges <,>
B: But the facilities is also there *na* <,>
A: Yeah <,> (ICE-IND: S1A-054#137–140)

75. Cf. Lange (2007: 95f.) for an overview of the syntax and semantics of *itself* and other intensifying *self*-forms in Present Day English.

If we look at speaker A's first utterance in isolation, the first interpretation that suggests itself clearly is that the students are "being charged *only* that much (and not more)" for their hostel, i.e. alternatives which are higher on the scale of hostel fees are excluded. But the next clause cancels this default interpretation: the students are actually paying higher hostel fees than others. *Only* is this case is then unambiguously a presentational focus marker which serves to intensify what is in its focus: "we are being charged **that much**!"

Example (4.6.10) below represents an ambiguous case where the conversation contained examples of both restrictive *only* and presentational *only* and where the token in question admitted both interpretations. As in all previous chapters, I chose a conservative approach and did not include such ambiguous examples in the tagging for the presentational use.

(4.6.10) A: If <,> you are you are <.> goi </.> going to give a party <,>
C: Not so grand but light <,>
A: Uh <,>
C: Not so grand but light one <,>
A: Light one <,>
B: Light party <,>
A: Whom you are calling
B: In uh at house or in hotel
C: All friends <,>
B: At house **only** or at hotel <,>
C: At house **only** <,> (ICE-IND: S1A-051#198–208)

The distribution and functions of *only* and *itself* in the direct conversation files of ICE-India and ICE-GB are then depicted in Tables 4.32 and 4.33.

Table 4.32. Frequency of *itself* and *only* in the direct conversation files in ICE-India and ICE-GB

ICE-texts S1A 1–100	ICE-India (absolute/per 10,000 words)	ICE-GB (absolute/per 10,000 words)
itself	49/2.24	6/0.29
reflexive	2/0.09	1/0.05
intensifying	17/0.78	5/0.24
presentational focus	29/1.33	–
unclear	1/0.05	–
only	617/28.23	204/9.92
presentational focus	105/4.8	–
contrastive focus/other*	512/23.43	204/9.92

* This category might include the occasional token of *only* as a discourse marker, as in "the flowers are lovely; only, they have no scent" (Brinton 1998).

Table 4.33. Different functions of *itself* and *only* in the direct conversation files in ICE-India and ICE-GB, as percentage of the sum of all tokens

ICE-texts S1A 1–100	ICE-India	ICE-GB
itself	49/100	6/100
reflexive	2/4.08	1/16.66
intensifying	17/34.7	5/83.33
presentational focus	29/59.18	–
unclear	1/2.04	–
only	617/100	204/100
presentational focus	105/17.01	–
contrastive focus/other	512/82.98	204/100

Both *itself* and *only* (regardless of the actual meaning) are more frequent in IndE than in BrE; the difference in usage is most striking with respect to *itself,* which occurs more than seven times as frequently in IndE as compared to BrE. *Only* in general is almost three times as frequent in ICE-India, and *only* as contrastive focus marker still occurs more than two times as frequently in ICE-India in comparison to ICE-GB.

If we combine the figures for *it*-clefts from the preceding chapter with those for *only* as contrastive or exclusive focus marker, we arrive at the relative frequencies of syntactic and morphological contrastive focus constructions in the two varieties as depicted in Table 4.34.

Assuming that the necessity to express contrastive focus is more or less the same for speakers of IndE and BrE, then the statistics are surprising: the frequency for the two types of contrastive focus constructions combined is almost twice as high for ICE-India compared to ICE-GB. We might, with all due caution, speculate that the prevalence of morphological focus marking with *only* in ICE-India is a compensatory strategy: I have quoted Platt et al.'s observation in the chapter on dislocation (cf. 4.4 above) that the New Englishes are generally considered to make less use of intonation, i.e. of phonological means to indicate focus. A speaker of BrE

Table 4.34. Syntactic and morphological contrastive focus constructions in the direct conversation files of ICE-India and ICE-GB

Contrastive focus construction	ICE-India (absolute/per 10,000 words)	ICE-GB (absolute/per 10,000 words)
Clefts	7/0.32	46/2.23
Only	512/23.43	204/9.92
Total	519/23.74	250/12.15

would then be more likely to mark a contrastive focus construction by prosodic prominence than a speaker of IndE, who is more likely to prefer the free morpheme *only* as focus marker. I will return to the correlation between morphological and prosodic focus marking below.

4.6.1.3 Only *and* itself *as presentational focus markers in ICE-India.* As already said, IndE shares with other varieties of English the use of *only* as marker of contrastive focus, but has developed a new use for *only* as presentational focus marker, with *itself* occurring as an alternative to presentational *only*. While *only* is clearly much more common than *itself*, the two forms seem to appear in quite similar environments (cf. Lange 2007: 107). Both forms are frequently associated with locative or temporal foci:

(4.6.11) A: Last year also uh <,> I did one refresher course <,> in the month of June **itself** <,,> and uh the duration of that was was <,> twenty-four days (ICE-IND: S1A-075#167:1:A)

(4.6.12) A: Somehow like *no* it's very hot
Like in the afternoon **only** it is burning
The skin is burning like <,> whereas in Goa we have the cool climate (ICE-IND: S1A-001#18:1:A>

(4.6.13) A: Where is she staying <,,>
B: Mangalore <,,>
A: Mangalore **only**
B: Yeah <,,> (ICE-IND: S1A-044#68–71)

(4.6.14) B: The eldest one <,> he is uh <,> in the third year of Engineering
A: Oh <,>
B: He is doing instrumentation <,>
A: Where sir <,>
B: In Shahada
A: Shahada **itself** <,>
B: Uh he got uh <,> medal <,> in the first year
(ICE-IND: S1A-067#223–229)[76]

With only 29 examples of presentational *itself* in the direct conversation files, all generalizations about distribution are tentative. However, a comparison of all tokens of *only* and *itself* in the complete ICE-India corpus revealed two main differences in the distribution of the two forms: unlike *only*, *itself* does not interact with verbal foci, and it does not occur in negative contexts (cf. Lange 2007: 109f.). Example (4.6.15) is an illustration of both these contexts for *only*:

76. *Shahada*: town in Maharashtra where speaker B lives.

(4.6.15) A: So what Raju had thought <,> uh he has thought that I'll come with you <,> and what they had planned that they will not talk to me **only** <,> at all <,> (ICE-IND: S1A-052#108)

The polysemy of *itself* might inhibit the occurrence as postverbal presentational focus marker: *itself* in postverbal position is probably too closely related to the reflexive marking function to acquire a new meaning (Lange 2007: 110). *Only* in combination with negation adds particular emphasis to the negated proposition, which is also explicitly expressed in example (4.6.15): *they will not talk to me* ONLY <,> ***at all***.

There is further a fixed expression, *like that only*, which has no equivalent with *itself*:

(4.6.16) A: Yeah hostels are **like that only** *na* <,> (ICE-IND: S1A-049#100:1:A)

(4.6.17) A: I can also hear Swami from your office
B: Yeah yeah Swami talks **like that only** <,>
A: [Laughs] don't need a loud speaker <,>
(ICE-IND: S1A-098#106–108)

The data in Lange (2007) show that the occurrence of *only* and *itself* in their innovative usage is correlated with medium: *only* is very common in spoken language, but almost entirely absent from written IndE. By contrast, focus *itself* is much rarer, but seems to have been accepted as suitable for the written language. We are here therefore on the track of an IndE innovation that might be a candidate for inclusion in an emerging IndE standard – a consideration that will be given more space in Chapter 5.

4.6.1.4 *Explanatory parameters.* The ICE-India data on presentational *only* do not replicate Mesthrie's findings for SAIE. Recall that in this variety of IndE, *only* was first of all found to be restricted to basilectal speech. Secondly, *only* was exclusively used by speakers whose first language belonged to the Dravidian language family. The data for the distribution of *only* in correlation to the ICE-India speaker variables paints a rather different picture. Figures 4.35 – 4.38 depict the relative frequencies of both *itself* and *only* according to the familiar parameters gender, age, genetic affiliation of mother tongue and educational level. The low number of *itself*-tokens does not allow statistical significance testing, and all further observations should therefore be taken as no more than tendencies which could only be confirmed on the basis of a much larger corpus.

Figure 4.35 shows that gender is a relevant factor for the occurrence of both *itself* and *only*, with women exceeding the expected frequencies and men remaining below them. For presentational *only*, the distribution according to gender is highly significant (at $p = 0.001$). I will return to this point in the next chapter on invariant

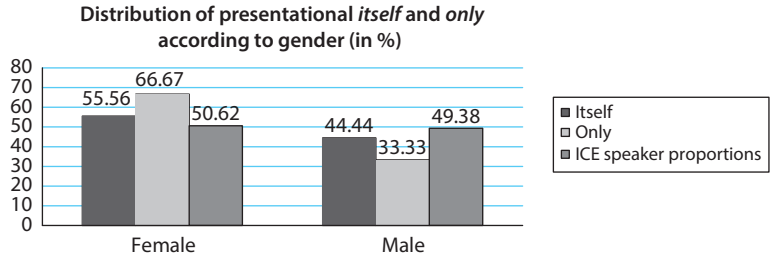

Figure 4.35. Distribution of presentational *itself* and *only* in the ICE-India conversation files according to speaker's gender

tags in IndE, which have also been claimed to occur more frequently in women's than in men's conversations (cf. Valentine 1991).

The next diagram indicates the tendency for both *itself* and *only* to cluster in the two youngest age groups. The reason for this may be the degree of formality: many of the conversations involving younger speakers take place between friends or within the family, whereas none of the speakers who are 42 or older are involved in conversation about, say, movies or hanging out with friends. The preference of younger speakers for the innovative use of *only* may also indicate ongoing language change, clearly a topic for further research.

The next diagram brings us back to Mesthrie's observations on SAIE. Dravidian speakers are overrepresented with respect to *only*, but the form is also well represented with IndE speakers whose mother tongue belongs to the Indo-European language family. Again, the fact that *only* does not occur in the speech of speakers whose first language belongs to the other language families may be noted in passing, but cannot be investigated further with the limited data available.[77]

The more or less balanced proportion of Dravidian and Indo-European mother tongue speakers with respect to *only* is much less surprising than Mesthrie's result, as the discussion below on the source of the form will reveal.

Finally, presentational *only* is definitely not restricted to basilectal varieties of IndE, as the mere fact of the form's presence in the ICE-India conversation already indicates. The figures for both presentational focus markers represent a rather mixed picture. We find *only* underrepresented with speakers who hold a PhD or a Master's degree and overrepresented with the two remaining groups, but still firmly entrenched within the ICE-India speech community. *Itself* does not pattern as evenly, though; the form is overrepresented with graduates and those with a secondary education. If more data were available, a cross-tabulation with the results for distribution according to age might provide deeper insights.

77. Only five speakers of Tibeto-Burman languages and one speaker of an Austro-Asiatic language are represented in the ICE-India direct conversation files, cf. Chapter 4.1.3.3.

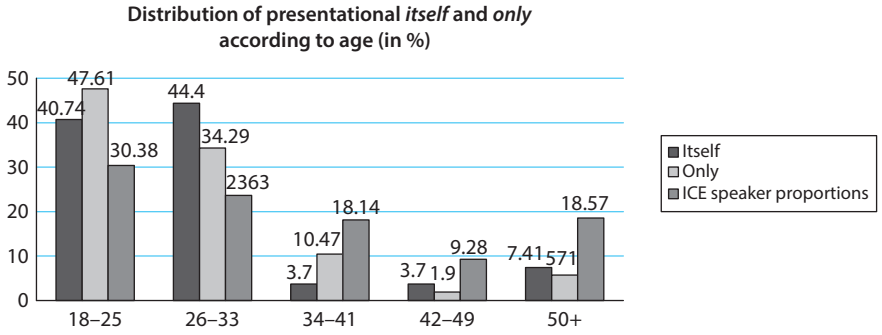

Figure 4.36. Distribution of presentational *itself* and *only* in the ICE-India conversation files according to speaker's age

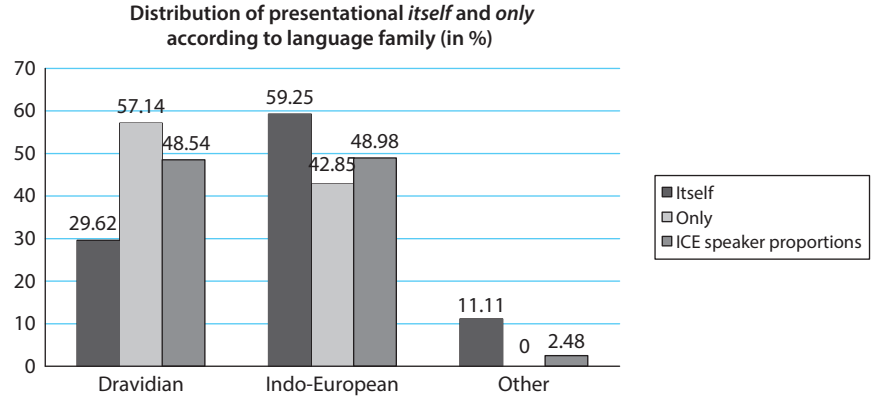

Figure 4.37. Distribution of presentational *itself* and *only* in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

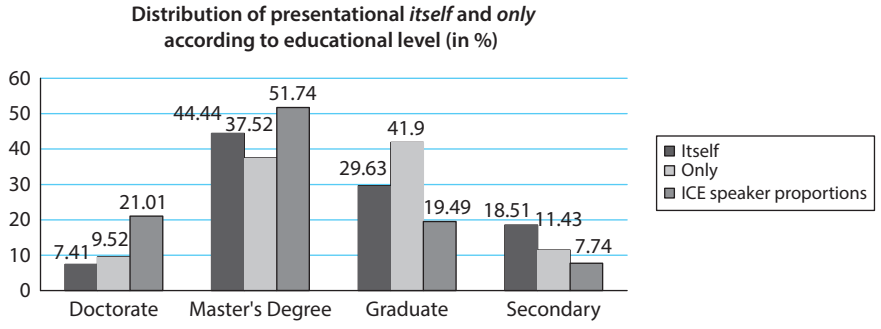

Figure 4.38. Distribution of presentational *itself* and *only* in the ICE-India conversation files according to speaker's educational level

It is uncontroversial that the innovative paradigm of presentational focus marking in IndE is contact-induced. Enclitic focus markers are widespread in Indian languages, as Masica confirms for the Indo-European language family (1993: 396):

> A feature of NIA [New Indo-Aryan] syntax that must also be kept in mind, however, is the set of *emphatic particles* (H. *hī*, B. *i*, G. *j*, M. *ts*, etc.) which make it possible to express "emphasis" without the help of either word order variation or intonation (although also not incompatible with either).[78]

Similarly, Krishnamurti traces the emphatic clitic *-ē to the clitics that can be reconstructed for Proto-Dravidian (2003: 413): "This clitic adds emphasis to the meaning of any constituent of a clause to which it is attached, broadly meaning 'only'" (2003: 415). Clearly, then, *itself* and *only* are calques and thus instances of pattern replication. The question then is where the presentational meaning of *only* and *itself* comes from, given that the corresponding form in Indian languages is generally described to have exclusive/restrictive meaning.[79]

Hindi, to take one example, has a three-way paradigm of optional "emphatic particles" (Agnihotri 2007a: 149) which can occur as enclitics to a wide range of constituents. *Hii* is generally paraphrased as 'only' and marks restrictive focus, *bhii* 'also' is an inclusive focus marker, and "*to* has a variety of meanings including, among others, 'then', 'so what', and 'so far as X is concerned'" (*ibid.*), that is, *to* is a topic marker. A more detailed description of the range of functions of the Hindi clitics is provided by Koul (1990). His examples for "different shades of meanings of the particle *hii*" (1990: 32) are instructive:

(4.6.18) a. mohan jaayegaa
'Mohan will go'

b. mohan jaayegaa **hii**
'Mohan will certainly go'

(4.6.19) a. mɛŋ gayaa nahiiŋ
'I did not go'

b. mɛŋ gayaa **hii** nahiiŋ
'I did not go at all'

(4.6.20) a. voh aaj gayaa hogaa
'He might have gone today'

b. voh aaj **hii** gayaa hogaa
'He might have gone today itself'

78. H.: Hindi; B.: Bengali; G.: Gujarati, M.: Marathi.

79. Cf. e.g. Sharma 2003 for Hindi.

(4.6.21) a. yah acchaa huaa
'It is good'

b. yah acchaa huaa **hii** huaa
'It is good (emphatic)' (*ibid.*)

These examples show that the Hindi particle *hii* already displays the same polysemy as *only* in IndE, marking both exclusive/contrastive as well as presentational focus. Example (4.6.19b) recalls example (4.6.15) above: when the speaker says *they will not talk to me* ***only*** *<,> at all*, she is replicating a pan-Indian pattern of focus marking. In this case, the postverbal clitic of the source language (in this case Kannada, a Dravidian language) is replicated twice, first as the already conventionalized *only* and then as *at all* to reinforce the emphatic meaning in a negative context. Example (4.6.20b) shows that *itself* is another target for replicating the Indian pattern of focus marking. The fact that both *only* and *itself* are chosen as the IndE equivalent of an enclitic focus marker derived from an Indian source language is not surprising. Crosslinguistically, the domains of reflexivity, intensification and focus marking are closely related (cf. König & Siemund 1999). In German, the form *selbst* can be both an intensifier and a focus marker, while a different form is used for expressing reflexivity:

(4.6.22) **Selbst** ICH habe das verstanden.
'**Even** I have understood that.'

(4.6.23) Ich habe das **selbst** gemacht.
'I've done that **myself**.'

English, on the other hand, has the same forms for marking reflexivity and intensification, but a different form, *even*, to mark the focus of a sentence. In Hindi, the enclitic form *hii* similarly participates in the domains of reflexivity and intensification. When the reflexive pronoun *aap* "is followed by *hii*, it has an adjectival intensifying force and qualifies a noun or a pronoun" (Koul 1990: 31); that is, *aap hii* has an intensifying function just like the English *self*-forms in e.g. *I have done it* ***myself*** or *The president* ***himself*** *will come to the meeting.* The question then is why *only* rather than *itself* is the preferred option that has stabilized as a feature of spoken IndE. Elsewhere, I suggested that

> *itself* occurs less frequently because of constraints from the source language: intensifying *itself* is rare in PDE and requires very specific semantic contexts, whereas *only* is much less restricted, both syntactically and semantically. *Only* is more variable concerning its position in the sentence, and it does not place any selectional restrictions on the focus. In other words, *only* is more flexible in its use, less dependent on specific contexts, and therefore probably easier to appropriate for new functions. Further, *itself* carries agreement features for person,

> number and gender, whereas focus particles such as *only* are typically uninflected. Although there are some instances in both corpora where *itself* is used as an invariant focus particle, it seems that syntactic constraints operating on the source item are not as easily "nativized" as semantic selectional restrictions: overt agreement violations are not a favoured option when *itself* is used as presentational focus marker. (Lange 2007: 114).

Matras & Sakel's concept of pivot-matching is again helpful to elucidate this particular area of linguistic convergence. In pivot-matching, "a construction is selected from the repertoire and replicated by re-constructing pivotal features around a new set of word-forms" (Matras 2009: 259). The Indian enclitic focus markers are calqued rather than borrowed, which is evidence for one of the few undisputed constraints on borrowing, namely that bound morphemes are comparatively impervious to borrowing (cf. Matras 2009: 155f.). The replicated construction in IndE preserves the structure of the Indian construction in that the focus marker generally follows its focus. The IndE speakers then have two options for re-creating the enclitic focus marker in the target language- both *only* and *itself* are available as loan translations, and both forms display the polysemy required for pivot-matching (cf. Chapter 2.2.3). However, *only* is clearly semantically more concrete than *itself* as well as syntactically less constrained, as shown above. Selecting *only* rather than *itself* is thus a paradigm case for the process of pivot-matching:

> Speakers will inevitably direct their attention to the more concrete meaning when searching for a match in the replica [...]. The match will then lead to the emergence of a more abstract meaning. (2009: 239)

I would claim that IndE presentational focus marking with *only* and *itself* does indeed represent a qualitative difference to other Englishes: an entirely new category is introduced into the syntax of English. I will return to this issue in the final chapter.

4.6.2 Invariant tags

4.6.2.1 *Overview.* Invariant *isn't it* has surfaced in almost all impressionistic descriptions of IndE ever since Kachru (1986, 1994) included the feature in his list of nativized IndE usages. Compare McArthur (2003: 322) for an example:

> The use of *isn't it* is common as a generalized question tag, particularly in the south: *They are coming tomorrow, isn't it? She knows him very well, isn't it?*

Agnihotri's slightly exasperated comment indicates that the topic might be more controversial than one would expect (1999: 191):

> The use of the invariant tag-question 'isn't it' or 'no' [...] is perhaps the most frequently cited feature of Indian English (often without an awareness that this is also a feature of several other varieties of English, e.g. Canadian English). Yet the pressures of Standard English are so strong that only 34 percent Delhites accepted 'You are coming, no?' as good English.

Whether "good English" or not, the form *no* has found its way into Indian English literature as a marker of the Indian "grammar of culture", as d'Souza claims (1991b: 312):

> What exactly does this ubiquitous 'no' do? It has some of the illocutionary force of the Standard English tag question, but it is more than that. It adds emphasis to the statement while at the same time establishing solidarity with the person spoken to. In most instances it does not call for an answer because it is taken for granted that the speaker and hearer share the same beliefs or sentiments. [...] The 'no' performs some of the functions of the Standard English tag question, but in addition it carries nuances of meaning that are special to the Indian variety of English. The 'no' should be seen as part of the Indian English system and not just as a deviation from the Standard tag. It has a paradigmatic relationship with the other English tags.

Indeed, the literature on IndE *no/isn't it* may well serve as an example for the development of research on IndE in general: it mirrors the fashions, preoccupations and ideologies prevalent in the last three decades of scholarly attention directed at the subject. Before I turn to a discussion of relevant contributions, I will briefly refer to Kortmann et al.'s (2004) evidence for the worldwide distribution of "invariant non-concord tags" (their feature 52). The next section, which is devoted to my corpus findings, will be introduced by an overview of the functions of tag questions in British English as a basis for comparison with IndE.

Invariant tags do not occur among the top features for the British Isles, America, the Caribbean, or the Pacific, but are prominent in Asian and African varieties of English (Kortmann & Szmrecsanyi 2004). Mesthrie regards invariant tags as "another overwhelming rule in L2 Englishes" (2008: 630), represented in six out of twelve African and four out of five South and Southeast Asian varieties.[80] This ranking among the top L2 features should be kept in mind when considering explanations that appeal to substrate influence: for a feature that is prevalent in speech communities on two continents, referring to substrate influence is bound to be too simplistic. However, if we follow Kachru in an early pronouncement on

80. Pakistani English is missing from Mesthrie's list of South Asian varieties which allow invariant tags, although Mahboob (2008: 584) lists the occurrence of this feature. It thus seems safe to say that invariant tags are common in all South and Southeast Asian Englishes.

invariant IndE tags, then invariant *isn't it* just represents the English calque of the Hindi negative particle *na*:[81]

> In English, the structure of tag-questions is composed of a statement and a tag attached to it. In such structures, there is contrasting polarity; a positive main clause is followed by a negative tag and vice versa. In Hindi-Urdu, the parallel structure consists of a single clause with a post-posed particle which is invariably *na*. Transfer thus results in South Asian English constructions such as *your* [sic] *are going there, isn't it?* and *he isn't going there, isn't it?* (Kachru 1986: 40)

A later, slightly reformulated version goes beyond the account above in adding a rather categorical statement about the distribution of *isn't it* in IndE and South Asian English generally:

> In British English the tag questions form a set, out of which an appropriate choice has to be made according to the context. In South Asian English generally, that choice is restricted to *isn't it*. *You are going tomorrow, isn't it?*, *He isn't going there, isn't it?* (Kachru 1994: 520)

Kachru seems to claim that invariant *isn't it* is the only available option for IndE tags, without offering any empirical evidence. For Sedlatschek (2007), it is this account of IndE tags which epitomises misguided research in the field of New Englishes. I will come back to Sedlatschek's criticism below; it is first worth noting that Kachru's treatment of invariant tags attracted unfavourable comments as early as 1985.[82] Sridhar's misgivings about Kachru's methodology are quoted here in full: first of all, Sridhar includes the *no*-tag as an alternative choice to invariant *isn't it*, thus going beyond Kachru who only referred to *no/na* as the Hindi origin of invariant *isn't it*. He further outlines a research programme that has only come to be taken up decades later:

> The contrastive [i. e. Kachru's] approach is useful only as an initial stage of analysis to identify the variables. It needs to be supplemented by (i) a ***systemic*** approach that seeks to explain how the putative features pattern into an autonomous system, and (ii) a ***sociolinguistic*** approach that seeks to predict the social and/or linguistic circumstances in which a given formal feature is likely to occur.
> I will briefly explain this point with selected examples of Indian English. [...] It has been claimed that lack of inversion of subject and auxiliary in yes-no questions, and the use of the invariant tag "isn't it" or "no?" in tag questions are two (of the many) important features of Indian English. However, there is quite a bit of variation in the frequency of occurrence of these features, with speaker's level of education being almost certainly the major determinant. (Sridhar 1985: 45, emphasis in the original)

81. *Na* and *no* are generally treated as equivalent, cf. below.

82. Kachru has been writing on this topic (and often re-writing his earlier articles) since the mid sixties.

What Sridhar seems to be insinuating is that high frequency of occurrence of invariant tags correlates with a lower degree of education: speakers with a higher degree of education (and, as a corollary, a longer and qualitatively superior exposure to English) may be more likely to use the standard variants than uneducated speakers. Another way of phrasing this hypothesis would be to claim that invariant tags belong to a mesolectal or even lower level of IndE, whereas acrolectal IndE favours the general English repertoire of tag questions.

Rogers subsumes both non-concord *isn't it* and *no* under the category "invariant tag" (2003: 55). She notes that both features are much more common in spoken as compared to written registers (which should not come as a surprise), but makes no attempt to explain speakers' choices in the matter. Hosali (2008: 573) indirectly supports Sridhar's assumption. She notes that the "tendency in IndE is to use *isn't it* as a universal question tag and *no* as one of its variant forms"; whereas in Butler English, a basilectal variety of IndE, *no*, *na* and *eh* are the only options available (*ibid.*).

The study by Sahgal & Agnihotri (1985) lends further support to Sridhar's hypothesis. The study elicited acceptability judgments from educated speakers of IndE on a prefabricated list of "instances of deviations which have been considered characteristic of IndE" (1985: 119), among them the invariant tags *no,* "so frequently heard in Delhi's colloquial English" (1985: 123), and *isn't it*:

> educated speakers evaluate the two forms differently. Whereas *no* is completely rejected, (2) "Your friend went home yesterday, isn't it?" is considered to be 'good English' by 25% and 'good enough for informal use' by another 39% of the informants. (*ibid.*)

The study highlights the phenomenon familiar from work on language attitudes: there is frequently a considerable gap between actual linguistic performance and acceptance, especially where spoken language is concerned. My data analysis below will address the question whether there is indeed a correlation between speakers' educational level and their choice of invariant tags.

Sahgal & Agnihotri (1985) also go some way towards meeting Sridhar's demand quoted above for "a sociolinguistic approach" in the study of IndE syntactic features, since their data can be used to "predict the social and/or linguistic circumstances in which a given formal feature is likely to occur" (Sridhar 1985: 45). However, Sridhar's call for "a systemic approach" is still not met: the evidence for the form and function of IndE invariant tags remains anecdotal. Schmied (1994) tried to provide some quantitative evidence for the Kachruvian feature list, using the Kolhapur Corpus of written IndE as the database. His remark on the occurrence of invariant *isn't it* in the corpus is instructive: "you tend to wonder what it means if a feature is listed as 'typical' and then it appears only once in a million word corpus" (1994: 223). Moreover, the sole token of *isn't it* that he found in the

corpus "occurs in a fiction text in direct speech" (*ibid.*).[83] A quick glance at the written part of ICE-India confirms Schmied's doubts about invariant *isn't it* as "typical": the search for *isn't it* turned up eight examples; five of these were tags, and only two of these five were invariant tags:

(4.6.24) The name of the place sounds like it is from some part of the African Continent **isn't it?** (ICE-IND: W1B-014#36:1)

(4.6.25) Tho' we think we have progressed but we have failed to admire nature **isn't it** (ICE-IND: W1A-002#109:2)

Example (4.6.24) comes from the text category "social letters", example (4.6.25) from the category "untimed student essays". Both belong to the superordinate category "non-printed" writing, a category that was absent from the Kolhapur Corpus.

It thus seems safe to conclude that invariant *isn't it* is a non-standard feature that has no place in the register of written/printed IndE, an observation that should come as no surprise to critics of "the Kachruvian enterprise" (cf. Chapter 2.1 above). The actual frequency, form and function of invariant tags in spoken IndE in general will be the topic of my next section. As noted above, I would like to conclude this overview with Sedlatschek's comments on the research history of IndE invariant tags (2007: 28):

> What is treated rather statically and superficially as a characteristic feature of educated South Asian English by Kachru 1994 is a linguistic phenomenon that may be common in colloquial IndE speech (further evidence would be needed to back up this claim) but one that is extremely rare in educated writing (at least as far as the texts genres represented by the Kolhapur Corpus are concerned). While *no* was unacceptable to a majority of educated users of IndE in the mid-1980s, *isn't it* was more acceptable and possibly on its way toward further integration. In the light of this interpretation, IndE looks different from what Kachru 1994 claims to be the case. It is not the restriction of choices in the use of tag questions that characterizes IndE but its variational profile across speech and writing, seeing the use of invariant tag questions in one domain of use and their absence in others.

4.6.2.2 *Tags in ICE-GB.* Tottie & Hoffmann's corpus-based study on "Tag questions in British and American English" (2006) is a useful point of departure if we want to consider the form and discourse functions of tags in L1 Englishes. Generally, tag questions are much more prominent in BrE than in AmE: in spontaneous conversation, tags are more than nine times as frequent in BrE than in AmE (2006: 288). Invariant tags such as *innit?* and *weren't it?* are exclusively confined to Tottie & Hoffmann's BrE data, and the low number of relevant tokens (27 out of a total of almost 5000 instances from the spoken part of the *British National Corpus*)

83. The Kolhapur Corpus was released in 1986 and comprised only printed material (from newspaper texts to fiction).

runs against Krug's claim (Krug 1998) that *innit?* is already on its way to entering the standard language (Tottie & Hoffmann 2006: 286). In the ICE-GB conversation files, there is only one *innit?* (tagged "UNTAG") among the 471 tokens that can be retrieved by searching for the functional label "TAGQ", a finding which clearly corresponds to Tottie & Hoffmann's figures for the rather marginal role of *innit?* in standard spoken British English.

Tottie & Hoffmann also looked at the possible combinations of pronoun and auxiliary in tags: there are differences between the two varieties in the ranking of the individual question tags, but *isn't it* is the most frequent form in both (20.4% in BrE and 18.6% in AmE). In BrE, *isn't it* is followed by *is it* (6.1%); "[a]ll remaining tags each account for 4 percent or less in both corpora" (2006: 296). *Isn't it*, then, would appear to be the prototypical member of the set of possible tag questions, an observation that might go some way towards explaining why the form is so frequently chosen as invariant tag in L2 Englishes.

For the analysis of the discourse functions served by tag questions, the authors make use of the classification depicted in Table 4.39. The first two categories, namely "informational" and "confirmatory", can be subsumed under the general heading "epistemic modal", since they comprise tags which indicate the speaker's attitude towards the content of his/her own utterance: the main function of these tags is to elicit information from the addressee. Informational tags are, however, rare in Tottie & Hoffman's corpus, accounting for only 4% of tokens in both varieties.

Table 4.39. Pragmatic functions of tags (adapted from Tottie & Hoffmann 2006: 300–302)

Pragmatic function	Definition	Frequency BrE	Frequency AmE
Informational	"genuine request for information"	4%	4%
Confirmatory	"the speaker is not sure of what s/he says, wants confirmation"	37%	30%
Facilitating	"the speaker is sure of the truth of what s/he says but wants to involve listener"	36%	50%
Attitudinal	"emphasizes what the speaker says, does not expect involvement or reply"	18%	12%
Peremptory	"immediately follows a statement of obvious or universal truth, with which it is practically impossible to disagree ... the speaker considers the conversation about it at an end [...] often a put-down of the addressee" (Algeo 1990: 447–48).	1%	1%
Aggressive	"functions as an insult or provocation"	1%	0%

Confirmatory tags, on the other hand, are the most frequent tag type in BrE (37%), e.g.:

(4.6.26) A: Oh but what type of property has he got
It's a flat **isn't it**
B: No
It's a three bedroomed <,> house with a <,> with an integral conservatory <,> (ICE-GB: S1A-061#233–236)

All other tags serve mainly "affective" functions of adding emphasis to the speaker's utterances, involving the listener in the conversation or even of challenging the listener.[84] Facilitating tags are almost as frequent as confirmatory tags in BrE (36%). The authors have "adopted the label *facilitating* for tags whose major function appears to make the addressee participate in interaction, although not always out of politeness" (2006: 300), e.g.:

(4.6.27) B: What time do the off-licences close <,>
A: I don't know
Could be too late <,>
Oh sods <,>
We haven't got any in store **have we** <,>
B: No
I've got a few cans (ICE-GB: S1A-039#190–196)

This definition departs from Holmes' use of the term "facilitative tag" and the difference between the two definitions captures the general difficulty in assigning unambiguous pragmatic functions to tags. For Holmes, facilitative tags "are examples of hedges which serve as positive politeness devices. They invite the addressee to contribute to the discourse" (Holmes 1995: 81).

It is this hedging function of tags which has been associated with women's powerless language in Robin Lakoff's *Language and Woman's Place* (1975), the founding text of Feminist Linguistics for better or worse. Lakoff claimed that women use far more tag questions than men, and that these tags indicate insecurity and hesitancy. Neither Lakoff herself nor any other researcher working on the topic was able to substantiate her claim in quantitative studies. One reason for this may be that the discourse meaning of tags is not fixed, but arises out of the situational context: "one person's feeble hedge is another person's perspicacious

84. Tottie & Hoffmann point out that the two main categories of tag questions are also marked off by different intonation patterns: following Holmes (1995), they associate epistemic modal tags with rising intonation and affective tags with falling intonation. Since information about intonation is not available for their corpus material, the classification above and their results "must therefore be regarded as very preliminary" (Tottie and Hoffmann 2006: 300).

qualification" (Talbot 2003: 474). Tottie & Hoffman have found slightly higher numbers of tags in women's than in men's speech in both BrE and AmE, but refrain from further investigation precisely because of the context-bound character of tags (cf. 2006: 304).

To return to Tottie & Hoffman's pragmatic classification of tags: the study reveals that "[t]hree types, *confirmatory, facilitating*, and *attitudinal* tags, together account for over 90 percent" (2006: 301) of both BrE and AmE tag questions. In both varieties, confirmatory and facilitating tags are much more frequent than attitudinal tags, although the proportions of the three types differ. As already stated, confirmatory and facilitating tags are almost equally represented in BrE (37% and 36%), attitudinal tags comprise the remaining 18% of tokens. By definition, these tags do not necessarily occur utterance-finally and initiate a new turn by another speaker, e.g.:

(4.6.28) I think perhaps having decided <,> uh we'd all made our separate decisions **hadn't we** and obviously had similar problems in deciding how we were going to go about this (ICE-GB: S1A-064#125)

The remaining two types of tags, namely the peremptory and the aggressive tag, only play a minor role in Tottie & Hoffmann's corpus findings. The difference between the two types relates mainly to the content of the preceding clause rather than the actual function, which is a verbal attack on the other participant(s) in the conversation, e.g.:

(4.6.29) B: And don't stick your hands under my nose like you usually do
A: Well I shan't be sitting next to you so I can't **can I**
B: All right I'm going (ICE-GB: S1A-068#121–123)

Example (4.6.29) is "aggressive" rather than "peremptory" because speaker A's statement is not one "of obvious or universal truth", but prompted by the situational context.[85] Tottie & Hoffmann decided to uphold the distinction between the two types in order to test the hypothesis that aggressive tags are prominent in BrE. They found that "the aggressive tag does in fact only occur in British English, but it accounts for a very low proportion of examples there", that is only four actual tokens out of 371 (1%) (2006: 301f.). In the following, I will attempt to apply the taxonomy of tags as described above to the set of tags available in spoken IndE.

4.6.2.3 *Tags in ICE-India.* The discussion above suggests that it would be very useful not only to consider invariant *isn't it*, but all tags available to the IndE speech community, i.e. the canonical tag with reversed polarity as in the examples from

85. I could not find an unambiguous example for a peremptory tag in ICE-GB.

ICE-GB above, the much publicised invariant tag *isn't it*,[86] and the particles *no* or *na*. In doing so I only coded those tags as "invariant" rather than "canonical" if the context was unambiguous, e.g.:

(4.6.30) No <,> the elections notification came on first or second November **isn't it?** (ICE-IND: S1A-100#184:1:B)

(4.6.31) Uh then uh <,> hey Rasika has come down **is it?** (ICE-IND: S1A-098#139:1:B)

"Canonical" tags, to follow Tottie & Hoffmann's terminology, consist of a main clause, the "anchor", and the tag which is realized by a sequence of auxiliary plus pronoun (2006: 283f.). The pronoun in the tag displays gender and number agreement with the subject in the preceding clause, and the auxiliary mirrors the main verb with respect to tense and number. Tags typically (but not necessarily, cf. *ibid.*) also have reversed polarity: a positive verb in the anchor triggers a negative tag and vice versa. The "canonical" versions of examples (4.6.30) and (4.6.31) above would then be:

(4.6.30) a. No <,> the elections notification came on first or second November **didn't it?**

(4.6.31) a. Uh then uh <,> hey Rasika has come down **hasn't she?**

The label "invariant" for the English-derived tags in IndE therefore refers to the combination of auxiliary and pronoun, not to the presence or absence of negation. So far, only *isn't it* has been associated with the notion "invariant tag", but *is it* occurs as well in ICE-India; I will return to the function and distribution of these variants below.

The same speaker who produced (4.6.30) above also uttered (4.6.32) below. This token was coded as "canonical" tag even if, as the discussion below will show, there is a strong tendency for complementary distribution of the two variants:

(4.6.32) Issuing the orders is the main thing **isn't it?** (ICE-IND: S1A-100#166:1:B)

Example (4.6.32) might be invariant and just accidentally display the same form as a reversed polarity tag. However, choosing any other than strictly syntactic criteria for assigning the *is it/isn't it* tags to one group or the other appeared unwarranted and speculative to me.

Na is marked as "<indig(enous)>" in the corpus and therefore easy to retrieve by concordancing. A search with WordSmith returned 197 tokens. The manual tagging changed the number of tokens for consideration somewhat: I excluded

86. Besides *isn't it, is it* also features as invariant tag.

two examples that were included in the dialogues, but assigned to a third party (generally marked 'X' and not listed as a speaker in the relevant dialogue) from the count. I further disregarded all instances of *na* that were part of a longer stretch of discourse in the mother tongue, e.g.:

(4.6.33) B: Anyway now he is acting smart because it is already due **na** <,>
A: *Accha accha accha* <,>
B: *Ekvees ko due hai* ***na*** (ICE-IND: S1A-094#385–7)[87]

Speaker B's first *na* is counted as a *na*-tag, but not the second, because it is embedded in the syntax of the source language and might be subject to different rules and usage restrictions. My classification finally came up with 188 tokens of *na*. *No* as tag is a different matter, at least for the ICE-India corpus compilers and transcribers: the high-frequency form is never marked as "<indig(enous)>" even though the context makes it abundantly clear that it is fully equivalent to *na*, as in the following examples from the same text and the same speaker:

(4.6.34) Nice movie *yaar* that song is there ***no*** *hai apna dil to awara* <,,> (ICE-IND: S1A-052#245#1:A)

(4.6.35) What happened you know that fellow is there ***na*** I told you that he'll be he will be staring at me in the college I wanted to show you that fellow <,,> (ICE-IND: S1A-052#260:1:A)

The speaker, a female student aged 18–25, is describing a movie to her friend (also a female student aged 18–25), referring to one of the movie's songs with the Hindi song title. A couple of turns later, the same speaker uses *na* in what seems to be an identical syntactic environment and with probably the same discourse function, but in this case the form is clearly marked as "<indig(enous)>" (cf. Lange 2009: 213). I have no explanation for the corpus transcribers' decision to treat *na* and *no* differently, especially as even a superficial glance at the conversation files strongly suggest that *no* also functions as a tag, an observation that is supported by the research referred to above.[88] Again, in order not to force any premature interpretation on the data, I have kept *no* and *na* separate for the time being. A detailed investigation of their distribution in the texts will show whether the two particles are indeed identical in function.

We thus arrive at a fourfold distinction of tags in IndE. Two of these, the "canonical" tag and the invariant tag, are common in other Englishes: the former is mostly associated with L1 Englishes, and the latter with L2 varieties. The other

87. Both A and B are speakers of Marathi, with Hindi as additional language.

88. Columbus (2010) found a difference in use between *no* and *na*; her arguments will be taken up below.

Table 4.40. Total occurrences of tags in the ICE-India conversation files

	English-derived tags		Indigenous tags		Total
	Canonical tag	Invariant *isn't it/is it*	*no*	*na*	
	18	68	424	188	698
	2.58%	9.74%	60.74%	26.93%	99.99%
No. of texts containing tag(s)	9	29	65	50	77
No. of speakers using tag(s)	9	36	115	74	152*

* The combined figure for all speakers who have tags in their repertoire is lower than the sum of the speakers using the individual tags because of overlap in usage.

group, again with two members, consists of invariant negative particles derived from Hindi as the source language.[89] Comparing the overall figures for the two sets is already quite instructive, especially in the light of Kachru's insistence on *isn't it* as a "typical" feature of IndE and Sahgal & Agnihotri's acceptability study (1985).

Within the domain of available options for tags, the English-derived forms are clearly in a minority. Invariant *is it/isn't it* does indeed occur more often than the general English form, but both taken together account only for a little over 12% of all tags in the conversation files.

Table 4.41 compares the overall distribution of tags in the Indian and the British ICE-subcorpus. Occurrences of tags cluster differently in the two corpora; on the one hand, tags are much more evenly distributed in the ICE-GB conversation files, with only seven files that do not contain any tags as against 23 tag-less files in ICE-India.

Table 4.41. Comparison of frequencies and distribution of tags in the ICE-India and ICE-GB conversation files

	All tags in ICE-India	All tags in ICE-GB
No. of tags	698	471
No. of texts containing tag(s)	77	93
No. of speakers using tag(s)	152 (62.8%)	180 (48.13%)
Average no. of tokens per text	9.06	5.06
Frequency per 10,000 words	31.94	22.9

89. Negative particles serving as invariant tags are not restricted to Hindi, cf. below.

On the other hand, the percentage of speakers who have tags in their repertoire is higher for ICE-India than for ICE-GB, and the average number of tag-tokens in those text which actually contain tags is equally higher for ICE-India.

The data in Table 4.40 imply that Kachru and many of those who took issue with him on this matter were collectively barking up the wrong tree: the prevalent issue to consider in the IndE system of tags is not the opposition between variable and invariant tags, but the fact that the preferred option for most speakers is an indigenous form. This is especially interesting when recalling the strong objections speakers raised against *no/na* in Sahgal & Agnihotri's study (1985): the study reported on judgments being made before 1985. In contrast, the ICE-corpus data were collected in the early 90s, and it seems highly unlikely that the stigma apparently attached to *no/na* had changed much in such a short time. *No/na*, then, is widely used by speakers of standard IndE, but not readily acknowledged as being part of the (spoken) standard. We might even be dealing with a case of "fusion" in Matras' sense, defined as "the wholesale non-separation of languages for both forms and functions of a given class of grammatical items" (Matras 2000b: 577). As mentioned above, Matras (1998, 2000a) has presented evidence that discourse markers such as tags belong to the category of "utterance modifiers", and that this category is highly likely to be affected by fusion in multilingual contexts. We will have to return to this aspect below.

Since invariant tags are a widespread feature of L2 Englishes across Asia and Africa, their occurrence may be motivated by universals of second language acquisition such as regularization and simplification (cf. Williams 1987). Another explanation that deals more specifically with the Indian context comes from Bhatt (1995, 2008, Mesthrie & Bhatt 2008). He favours a pragmatic account for invariant *isn't it*:[90]

> In fact, tags in vernacular IndE are a fascinating example of how linguistic form is constrained by cultural requirements of politeness. More specifically, these undifferentiated tags are governed by the politeness principle of non-imposition. They serve politeness functions, signalling deference and acquiescence. (Bhatt 2008: 552)

Invariant *isn't it* in IndE would then be a prime example "of new English linguistic forms exhibiting a different 'grammar of culture' [...] from StdEng" (Mesthrie & Bhatt 2008: 135). The concept "grammar of culture" for the South Asian sociolinguistic area has been discussed in Chapter 3.2.4 above. Mesthrie & Bhatt (*ibid.*) acknowledge Bright (1968) in passing as the source of the term, but do not mention d'Souza or give any further characterization of the notion.

90. Bhatt does not mention *no/na* as an alternative to invariant *isn't it*, nor does he consider invariant *is it*.

In the following, I will first offer a descriptive account of the distribution of English-derived tags in the corpus. In a next step, I will focus on the meaning of these tags to see whether Bhatt's theory can be sustained. I will then extend the analysis to the whole range of tags available to speakers of IndE, that is including indigenous *no/na*.

4.6.2.4 *English-derived canonical and invariant tags: Distribution, function and parametrical analysis.* As said above, the repertoire of English-derived tags comprises both canonical tags and two invariant forms, *isn't it* and *is it*.[91] There are only 18 canonical tags in the corpus, uttered by nine different speakers in nine different conversations. Six of these conversations have only one single relevant token, and the remaining twelve tokens, that is two thirds of the total figure, can be attributed to only three different speakers. A distribution such as this precludes any meaningful statistical analysis – however, if I do refer to speaker statistics in the following, these should only be taken to indicate some general trends and not as statistically sound statements. Table 4.42 gives an overview of the pragmatic functions of both canonical and invariant tags in the data.

More than two thirds of the canonical tags and around half of the invariant tags can be classified as 'facilitating', e.g.:

(4.6.36) B: We may say that ideals <,,> will cease to be ideals if they are attainable <,> **won't they** <,>
A: Yes it is really sir <,> it is really true (ICE-IND: S1A-066#31–32)

Table 4.42. Pragmatic functions of English-derived tags in the ICE-India conversation files

	Canonical tags (total/%)	**All invariant tags (total/%)**	**Invariant tags (excluding independent utterances) (total/%)**	**Invariant tag: *is it* (total/%)**	**Invariant tag: *isn't it* (total/%)**
Informational	–			–	–
Confirmatory	2 (11.11%)	8 (11.76%)	8 (14.54%)	5 (21.74%)	3 (6.66%)
Facilitating	11 (61.11%)	33 (48.52%)	33 (60%)	5 (21.74%)	28 (62.22%)
Attitudinal	2 (11.11%)	9 (13.23%)	9 (16.36%)	–	9 (20%)
Peremptory	1 (5.55%)	2 (2.94%)	2 (3.63%)	–	2 (4.44%)
Aggressive	2 (11.11%)	1 (1.47%)	1 (1.81%)	–	1 (2.22%)
Independent utterance	–	13 (19.11%)	–	12 (52.17%)	1 (2.22%)
Unclear	–	2 (2.94%)	2 (3.62%)	1 (4.34%)	1 (2.22%)
Total	18	68	55	23	45

91. There was one instance each of just *isn't* and *isn't this*, which were subsumed under invariant *isn't it*.

(4.6.37) B: You were in Siligudi **isn't it**
A: Yeah but we were forced to stay in Siligudi because my parents were working there (ICE-IND: S1A-014#80–81)

(4.6.38) B: What movie had you been
A: *Khalnayak*
B: You had seen it before **isn't it**
A: I had – saw – seen it twice you know (ICE-IND: S1A-038#40–43)

An example of a confirmatory invariant tag would be the following:

(4.6.39) B: You gave your address **is it?**
A: Yeah <,> (ICE-IND: S1A-098#43–44)

The presence or absence of negation in the invariant tags does not correlate with the form of the anchor clause: most of the anchor clauses for *is it* are positive as in example (4.6.39) above. Neither does the choice between *isn't it* and *is it* seem to correlate with a difference in function: both are "chunks", unanalysable wholes. Interestingly, more than half of all tokens of invariant *is it* are independent utterances, a label I have chosen to indicate that they are not "anchored" to some preceding clause uttered by the same speaker, e.g.:

(4.6.40) B: Actually I don't read Kannada novels
A: **Is it**
B: It is
A: Then what what novels do you read (ICE-IND: S1A-017#67–70)

(4.6.41) C: She is physically very fat <,,>
B: [Laughs]
I see <,> **is it**
C: No you haven't seen her <,,> (ICE-IND: S1A-091#214–217)

(4.6.42) A: Uh <,> I have published <,> a book or two
B: **Is it** <,>?
A: Yeah <,> I have written a book for <,> uh B A English major students
B: Ah ha that is something [one word] (ICE-IND: S1A-026#50–53)

These independent utterances which bear some surface similarity to tags do not fit easily into Tottie & Hoffmann's classification scheme for the pragmatic functions of tags (cf. Table 4.39). Their categories apply to the combination of anchor plus tag uttered by one single speaker, thus expressing either his/her epistemic stance or his/her attitude, but always with respect to their own utterance. The speakers who utter *it it?* as an independent utterance initiate a change of turn and an

immediately following answer to their question: *is it* in these examples may be paraphrased by *really?* or *is that so?*, a slightly incredulous question that prompts the other speaker to meet the conversational challenge and to confirm what he or she has stated before. The speaker who utters *is it?* seeks confirmation, but not for his or her proposition, but for the preceding speaker's proposition.

Bhatt's claim for a different "grammar of culture" being made manifest by the use of invariant tags in IndE rests on this set of contrasting examples (cf. Bhatt 1995: 253, 2008: 1022, Mesthrie & Bhatt 2008: 134):

Unassertive/Mitigated (IndE)	Assertive/Intensified (BrE/AmE)
(1) a. You said you'll do the job, *isn't it*?	(1) b. You said you'll do the job, didn't you?
(2) a. They said they will be here, *isn't it*?	(2) b. They said they will be here, didn't they?

The (a) examples are supposed to belong to the register of informal IndE, the (b) examples to the register of formal IndE (Mesthrie & Bhatt 2008: 132). Mesthrie & Bhatt then explain (2008: 134):

> In contrast to the (b) examples above, Ind[ian] Eng[lish] speakers find the (a) examples non-impositional and mitigating (Bhatt 1995), placing them high on a scale of politeness. [...] In a culture where verbal behaviour is severely constrained, to a large extent, by politeness regulations, where non-imposition is the essence of polite behaviour, it is not surprising that speakers of Ind Eng use undifferentiated tags to sound less demanding in the requests that they make. Such tags have the added advantage of avoiding pronoun forms for the addressee, which would require other deference strategies.

In order to translate Bhatt's claim into a testable hypothesis, we need to look more closely at his premises.

Bhatt adopts the classification of tags made by Algeo (1988) which corresponds largely to Tottie & Hoffmann's taxonomy in Table 4.39. Table 4.43 gives an overview of the two systems. The only difference that is more than merely terminological concerns Algeo's category of confirmatory tags, which comprises tags that are "used to invite agreement from the hearer" (Mesthrie & Bhatt 2008: 133). Tottie & Hoffman have given the label "facilitating" to this function of tags, reserving the notion "confirmatory" for tags whose primary meaning is epistemic (cf. Tottie & Hoffman 2006: 300 and Table 4.39).

Bhatt now divides Algeo's five-fold classification of tags along the line of polite vs. impolite tag expressions: informational and confirmatory tags "index politeness, whereas the last three tags signal impoliteness" (Mesthrie & Bhatt 2008: 133). If we combine the figures for informational, confirmatory, and facilitating tags

Table 4.43. Comparison of tag classification taxonomies

Tottie & Hoffmann (2006)	**Algeo (1988) (quoted after Mesthrie & Bhatt 2008: 133)**
informational	informational
confirmatory facilitating	confirmatory
attitudinal	punctuation
peremptory	peremptory
aggressive	aggressive

from Table 4.32 above, we arrive at 72.22% "polite" canonical tags and 60.28% "polite" invariant tags: contrary to Bhatt's theory, the canonical tags tend to be used in more "polite" contexts than invariant tags. Further problems arise when we go beyond the rather coarse dichotomy of "polite" vs. "impolite" tags suggested by Bhatt. Attitudinal canonical tags, for instance, are not necessarily impolite, as the example illustrates:

(4.6.43) A: No I am not interested in shopping whenever I go to market I buy books only that is all <,>
B: So you are a bibliophile [laughter] **aren't you** <,,>
So I would better ask you to go to the <,> Mahalaxmi temple <,> have you been there[92]
A: Yes I've been there (ICE-IND: S1A-066#134–137)

This excerpt comes from a conversation between two lecturers who are attending an English refresher course at the university in Kolhapur. Speaker B (male, 50+, M.A.) is asking whether speaker A (female, 26–33, Ph.D) is going to go shopping on a free day after the course, and this example seriously casts into doubt that "attitudinal" tags per se should be subsumed under "impolite".

A final attempt at saving the theory might focus only on the "aggressive" tags, where canonical tags outnumber invariant tags. Since there are only three relevant examples in the data, their distribution cannot be called upon to rescue Bhatt's theory, but it might be instructive to consider the relevant tags in their contexts. The two "aggressive" canonical tags both come from speaker A in file S1A-55, who alone contributed five canonical tags, the highest number for an individual speaker in the data. A is male, age 18–25, and acts as the interviewer in the conversation. He is obviously having a hard time being taken serious by speaker B (a female medical student, age 18–25) and keeping the conversation going:

92. *Mahalaxmi temple*: temple for the goddess Laxmi in Kolhapur, of great historical and religious interest.

(4.6.44) A: Have you just come from Belgaum
B: No I came yesterday evening actually <,>
A: What were you doing there for so long <„>
B: Studying and <,> my medicine <,>
A: Oh <,> what subject <,>
B: Asking what subject <,> [laughter]
A: I mean <,> you just said something **didn't you** <„>
You have holidays now **don't you** <,>
B: Yeah we got ten days holiday <,> (ICE-IND: S1A-055#10–18)

When B in her third turn quite clearly indicates that she considers the question about subjects, if not the whole interview situation, quite ludicrous, she prompts A's exasperated and slightly hostile response *you just said something **didn't you***, ending in one of the tags which I classified as "aggressive". With his next utterance, *You have holidays now **don't you***, speaker A tries to steer speaker B back into the conversation, prompting her to react with the help of the facilitating tag. The conversation, however, does not go smoothly – speaker B continues to laugh and to give silly answers, so that A admonishes her with *Now let's try and keep this conversation serious* (ICE-IND: S1A-055#31). Speaker B retorts with *what you are recording this for* (ICE-IND: S1A-055#37), which shows that she is highly conscious of the interview situation. The following excerpt contains A's second "aggressive" tag:

(4.6.45) A: So uhm <„> how often you watch T V <„>
B: Couple of hours a day <„>
A: And what are your favourite programmes <„>
B: Oh wrestling <,> [two words] <„>
A: And <„> do you watch star T V <,> star plus
B: Yes star plus <,> a few Indian movies <„>
A: Oh you are not here all the time to watch T V <,> **are you** <„>
Uhm I prefer star T V <„> I do much like Zee T V <„> and I watch <,> I watch mainly watch star plus <,> because I don't much like sports I don't much like uh <,> Chinese channel because I can't understand Chinese <„>
Uh so I mainly watch star plus <„>
We have ten minutes more to go so you better think of something to talk about <„> (ICE-IND: S1A-055#190–199)

A has abruptly changed the topic of the conversation to television, but B only gives minimal responses, triggering another angry response by A. He concludes his lengthier comments on his own TV habits with an open and quite hostile challenge to speaker B to take over more responsibility for the success of the conversation.

This conversation is surely the least cooperative among the data. There are others in which speakers get worked up over a specific topic and disagree heartily, but none in which a speaker so openly signals her refusal to cooperate in the ongoing conversation. I have quoted and discussed the preceding excerpts in such detail to drive home the point that given the overall context, it is rather surprising that speaker A does not use even more "aggressive" tags. The whole conversational setup is aggressive, rather than the tags themselves; they receive their interpretation from the context.

The only invariant "aggressive" tag occurs in the excerpt below, an exchange between speaker B, a female school teacher (age 18–25), and speaker A, a male high school teacher, age 26–33. B is freely venting her frustration about teaching mathematics at school:

(4.6.46) B: But how come we will be every time doing the same thing <,>
We have to complete our portion within time <,> **isn't it** <,,>?
And again hammering the same and same and same sums we should complete our portion also
We should have <,> all the revision also <,>
How is it possible sir <,,>?

A: So in only one week <,> daily we must do <,> one one sum <,> on that all exercise <,,>

B: Sir we are having only half an hour period sir how come its <,> uh possible to do <,> because in one week time we're having seven period of maths <,> **isn't it** <,>?

A: Uhn

B: And in seven period atleast we should complete two to three excercises <,,> in seven period <,>
One exercise <,> teaching and doing the excercises and their homework and their classwork <,> then assignments are also there <,>
And completing this much is impossible sir

(ICE-IND: S1A-087#96–106)[93]

I classified speaker B's first invariant tag as "attitudinal" and the second as possibly "aggressive", prompting a rather helpless *Uhn* from speaker A in reaction to B's heavy complaints. One might argue that this second token might reasonably be classified as "attitudinal" rather than aggressive as well, leaving us with only the canonical "aggressive" tags. This would, however, not really impinge upon my argument. It is the situational contexts of conversations which might generate

93. "Excercises" is obviously a misspelling here.

"aggressive" tags, and not the other way round: there are no tags which are more or less "aggressive" by virtue of their form.

Further, what Bhatt seems to suggest is that the choice between canonical and invariant tags is determined by the parameters "formal vs. informal" in addition to the parameter "impolite vs. polite". It is not altogether obvious how Bhatt's claim concerning the parameter of (in)formality may be tested on our data. The fact that invariant tags clearly outnumber canonical tags is in itself not conclusive, as this preference could have any number of reasons. From the available evidence, I would be very hesitant to conjecture that the choice of canonical vs. invariant tags is dependent on the formality of the utterance context. First of all, all ICE-India private dialogue files generally represent quite informal discourse. The preference for a specific type of tag is therefore more likely to depend on the individual speaker, either because of his/her sociolinguistic profile or his/her personal style.

The diagrams 4.44 to 4.47 depict the speaker variables for the two English-derived tags. As already mentioned, the token frequency for canonical tags is not only very low, the distribution of the tokens is also highly skewed. Table 4.41 showed that the 18 canonical tag tokens occurred in the speech of nine speakers. Within this rather small group, three speakers stand out because they use the canonical tag repeatedly and taken together provide two thirds of all tokens.[94] The remaining six speakers only contribute one token each. Concentrating on the distribution of the slightly more numerous invariant tags in the ICE-India speech community, we find women to be overrepresented in their use of invariant tags.

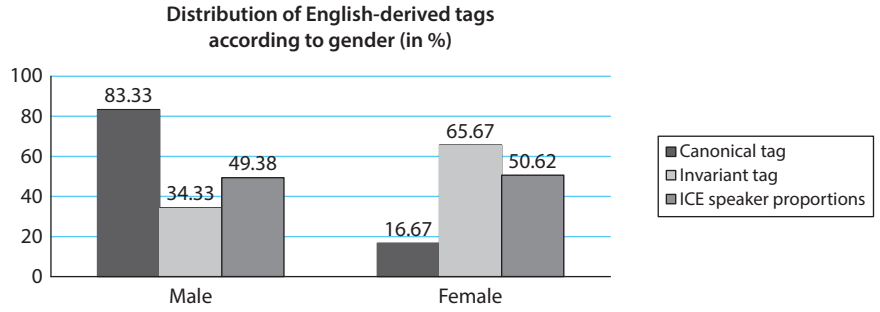

Figure 4.44. Distribution of canonical and invariant tags in the ICE-India conversation files according to speaker's gender

94. Speaker A in text S1A-55 (five tokens) is male, 18–25, graduate, mother tongue Marathi; speaker B in text S1A-66 (male, 50+, Master's degree, mother tongue Gujarati) contributed four tokens, and speaker B in text S1A-100 (male, 34–41, PhD, mother tongue Telugu) is responsible for three tokens.

The two youngest age groups show a proportionally higher preference for the invariant tag than the remaining three older age groups. A similar picture pertains to the variable "educational level": invariant tags are underrepresented with speakers who hold a PhD or a Master's degree and overrepresented with graduates and those with a secondary education.

It is particularly unfortunate that we do not have enough data to test Sridhar's assumption quoted above, namely that education is the most important factor for the choice between the invariant and the canonical tag. After all, the absolute figures for the canonical tag in Table 4.46 are only seven tokens for speakers with a PhD, five for speakers with a Master's degree and six for graduates, respectively. The slight overrepresentation of invariant tags with speakers of Dravidian languages is also not sufficient evidence to confirm McArthur's quote from the beginning of this chapter, which allocated the invariant tag mainly to South India.

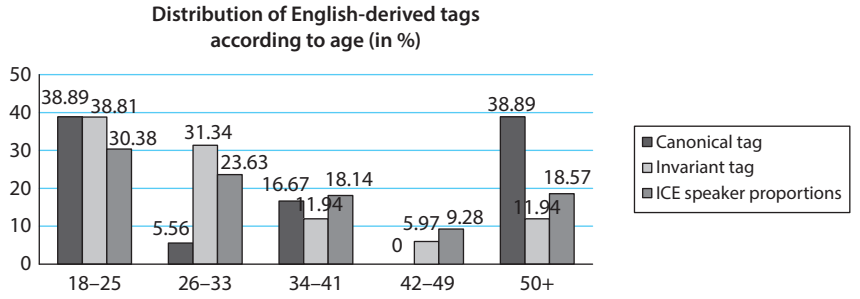

Figure 4.45. Distribution of canonical and invariant tags in the ICE-India conversation files according to speaker's age

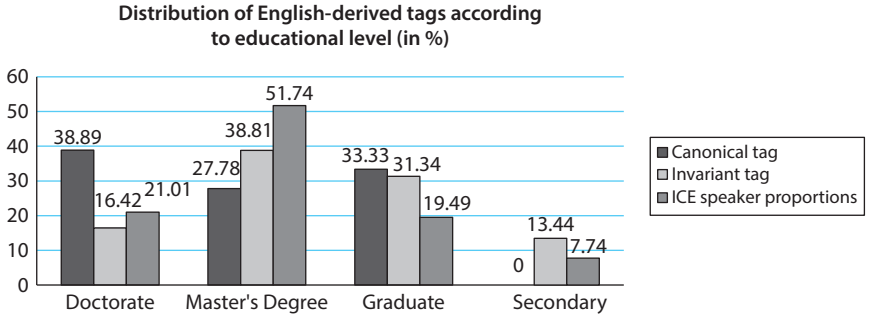

Figure 4.46. Distribution of canonical and invariant tags in the ICE-India conversation files according to speaker's educational level

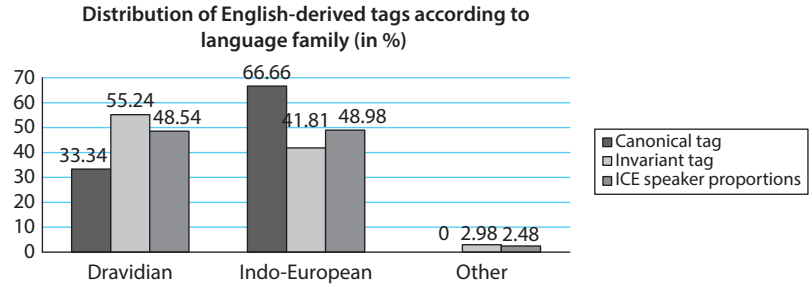

Figure 4.47. Distribution of canonical and invariant tags in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

Circumstantial evidence for the hypothesis that the choice of canonical vs. invariant tags depends on the individual speaker's style comes from the distribution of the tags in the corpus. Invariant tags occur in 29 texts, of which only five also contain canonical tags. Of these five texts, four have only one token of a canonical tag that may or may not be just accidentally canonical. Text 98, for example, has six invariant tags and one that I coded as canonical tag, but which might as well be invariant, given the strong preference speakers display for the invariant form:

(4.6.47) A: How do you like it?
Interesting **isn't it**?
B: Good <,> (ICE-IND: S1A-098#225–227)

My preliminary conclusion is, then, that the two English-derived tags are largely in complementary distribution, depending not on the speech situation or the degree of politeness to be expressed, but on speakers' personal styles. The only text containing invariant tags which also has more than one stray example of a canonical tag is text 100 (a telephone conversation), with four invariant and three canonical tags, all from the same speaker. Assuming that the canonical tags in text 100 are indeed canonical, text 100 can be used to see whether the speaker prefers invariant tags over canonical tags in those contexts that are deemed impolite, that is with attitudinal/punctuational, peremptory and aggressive tags. I classified six of speaker B's tags as "facilitating" and one as "confirmatory":

(4.6.48) B: Nizamabad comes under Osmania **isn't it?**
A: Yeah yeah <,> (ICE-IND: S1A-100#240–241)[95]

95. The utterance refers to constituencies in an election in Andhra Pradesh.

(4.6.49) A: Sir <,> from Monday onwards I too <,> want to take leave sir for four days <,> because total I have five CLs <,> so from Monday to <,> twenty ninth <,>
B: Uhm
Monday is a holiday function day **isn't it?**
A: Oh Monday is a holiday <,>
B: Maybe (ICE-IND: S1A-100#295–299)

Both tag categories belong to the "polite" usage of tags, so that there is no evidence for alternation between canonical and invariant tags conditioned by degrees of politeness in Bhatt's sense. Of course, many more texts such as this one would be needed.

So, unfortunately, this matter must remain unresolved until much larger corpora of spoken IndE become available – and until more empirically grounded hypotheses are put forward, I venture to add. After all, Bhatt's account of invariant IndE tags is open to the same charges made against Kachru: Bhatt's original observations on the topic were not backed up by any empirical evidence whatsoever, as his examples for invariant tags expressing politeness were constructed in order to contrast with textbook examples of BrE tags. Returning to his point in subsequent articles, he presented his original conjectures as an established fact, without ever having provided data from actual IndE language use. As valuable as his insistence on recognizing a different "grammar of culture" at work in IndE may be, his cause is ultimately not served well by following Kachru's example of combining far-reaching claims with very little sound evidence. However, this debate will ultimately prove to be a minor issue, since English-derived tags are numerically not as prominent as the indigenous tags, to which I will now turn.

4.6.2.5 *Indigenous* no/na*-tags: Distribution and function.* Since the indigenous tag *no/na* is derived from Hindi, it is convenient to first consider its functions in the source language before turning to the IndE data. One of the questions that has to be addressed in this context is whether we are dealing with borrowing or code-switching when speakers use *no/na* rather than an English-derived tag. The distinction between matter replication or borrowing and code-switching has been extensively discussed (cf. Chapter 2.3), and I will come back to this question below after treating the distribution and functions of *no/na* in greater detail.
Na belongs to the three-way paradigm of negative particles in Hindi (cf. Kachru 2006: 180–186) and can also be used to form tag questions. This multifunctionality is not restricted to Hindi, but is "a typical Indo-Aryan device" (Montaut 2005: 96). There are several options in Hindi for forming tags: *na* alone or in combination with other elements of the clause:

> The tag may contain a copy of the entire verb phrase minus the nominal elements, if any, followed by the negative particle [...] or just the tense marker [...]. Such questions are leading questions, with a strong expectation of agreement. (Y. Kachru 2006: 183)

Y. Kachru further provides some examples for *no/na* as a discourse marker, with the functions of "inviting agreement" and "signalling entreaty" (2006: 269f.). Her examples illustrate each discourse function:

(4.6.50) A: əb to ap xuʃ hɛ̃ nə?
now PTCL you.HON happy BE.PRES.PL TAG
'At least now you are happy, aren't you?

B: hã.
'Yes.'

(4.6.51) pita jī, mere liye yəh kʰilɔna kʰərīd dījiye nə.
father HON. I.GEN.OBL. for this toy.M.SG. buy give.HON.IMP. TAG
'Father, please buy me this toy.'

Similarly, Agnihotri notes that clause-final *na* in Hindi indicates "persuasion and request" (2007a: 23); it is used "to seek confirmation of or reaction to what has been stated earlier" (2007: 30). Masica confirms that most "NIA [New Indo-Aryan] languages also make use of (invariant) tag questions ('Isn't it'?) of the type of Japanese *ne* – in fact, in Gujarati the particle is also *ne*? (in Hindi it is *na*?)" (1993: 389). The question tag in Marathi is also *na*, e.g.:

(4.6.52) to gharī gelā nāhi, na?
he home-LOC go-PST-3SM NEG TAG
'He did not go home, did he?' (Pandharipande 1997: 40).

These negative particles do double duty for sentence negation. In this respect, Tamil and the other Dravidian languages are notably different from Hindi and other NIA languages: negation is expressed by negative verbal suffixes and negative lexical or auxiliary verbs rather than negative particles.[96] However, for tag questions, a strategy comparable to Hindi is followed, as Schiffman reports for Kannada (1983: 109):

> In tag questions, speakers follow a declarative sentence with the equational negative particle *alla* [...] plus interrogative *-aa*: *allavaa*. This is used to verify all kinds of propositions, equational and others.
> *niiv bandidri, alva?* 'you came, didn't you?'

96. See Lehmann (1993: 228–231) for Tamil, Krishnamurti (2003: 348–357) for Dravidian in general.

> *niiv shivrao, alvaa?* 'you are Shivarao, aren't you?'
> Unlike English, where the verb or auxiliary must be repeated, with a pronoun, in Kannada *allavaa* is used with any utterance, even if the original utterance is negative.

These accounts of invariant particles being employed as tags are remarkably similar to the pragmatic functions associated with the English tag questions. Moreover, question tags in Indian languages also appear to share the sentence-final position with their English counterparts. The functions for *no/na* mentioned above can be mapped onto the tag functions as labelled by Tottie & Hoffmann: seeking confirmation or some kind of reaction to a previously established proposition corresponds to the "confirmatory" and "facilitating" use of tags. It seems that these two most prominent functions of BrE tags, which account for over 70% of all tags in Tottie & Hoffmann's study, are also the most important functions of the Hindi *na*-particle. Kachru's example in (4.6.50) above could also be interpreted as "attitudinal", but it would take more context and a native speaker of Hindi to provide a definitive judgment on this. Where the functions of *no/na* and the canonical English tag question do not match is in the domain of what Kachru called "signalling entreaty": in her example quoted as (4.6.53) above, the *na*-tag follows an imperative. Agnihotri provides a similar example, originally intended to exemplify that the postposed *na*-particle does not convey a negative meaning: whereas *na bolo* means "don't speak", *bolo na* means "please say something" (2007a: 21). In this "entreating" function, the imperative in combination with the tag serves to elicit some action from the hearer in favour of the listener.

None of the available grammatical descriptions mentions usages akin to the "peremptory" and the "aggressive" use of English tags, and it will be the task of the next section to see whether *no/na* in IndE may also take over these functions. Previous studies which deal with *no/na* in IndE come from Valentine (1988, 1991) and Columbus (2010). Valentine found that *na* "has the interactive function of including the listener, which question forms usually do" (1988: 148). In a footnote, she adds that "these markers used in final utterance position gain a response [...] 100% of the time. When it is used medially, strictly being used as a maintenance device, it still evokes a response" (1988: 157). Her summary on the function of *no* runs as follows:

> What is commonly referred to as the 'no' tag is the use of the Hindi particle *na* attached to a positive or negative statement. The tag is often treated as an uncomplicated question form which functions as a device to elicit a response from the addressee. However, a tag may serve more than one function in terms of its affective meaning. Tag forms may express solidarity toward the addressee,

> encourage participation in the conversation, or express politeness. [....] What is interesting about this phenomenon [the transfer of the Hindi tag into the English discourse] is that the tag is not really functioning in the sense that the speaker is asking a question and expecting a response from the listener. Rather, is [sic] functioning as an interactive strategy strictly to make contact with her listener. (Valentine 1991: 331)

Valentine stresses the "affective" functions of *na*-tags and denies, or at least plays down, the "epistemic modal" function. Her observations on the discourse context and function of *no/na* in IndE will be compared with the evidence from the ICE-India data.

Columbus' study of invariant tags in several varieties of English is based on the ICE-corpora and includes *no/na* as the most frequent invariant tag in IndE. According to her (2010: 307),

> What is striking about *na* is that it is almost a synonym for *no* as a tag, but can be used for affirmation despite the origin as a negative particle. There is only one meaning for *na* which is not shared by *no*, that is for affirmation or confirmation of a previous statement [...]. Since the process requires complete semantic bleaching to occur, this could be an area for further acceptability and/or psycholinguistic research. Furthermore, IndE seems to prefer the use of *no* almost twice as much. It seems, then, that unless the meaning intended is to confirm the previous statement in some way, the use of *no* is more likely than *na* as an invariant tag.

However, neither my overall figures for the available *na-* and *no-* tokens in ICE-India (which differ from hers) nor my classification of their functions seem to support her analysis. Columbus differentiates 19 different functions for tags (cf. 2010: 298), which are not readily comparable to the taxonomy I adopted from Tottie & Hoffmann (2006). Moreover, I found no evidence for her claim that *no* and *na* behave differently with respect to the confirmation/confirmatory function (cf. Table 4.48). Such discrepancies are bound to arise with highly context-bound forms such as discourse markers, whose interpretation invariably involve subjective judgments.

To summarize: English tag questions and the Hindi particle *no/na* display a striking functional isomorphism. Both forms express affective as well as epistemic modal meanings in discourse, but possibly to different degrees. The quantitative and qualitative analysis of *no/na* as it appears in the ICE-India conversations yields Table 4.48. For ease of comparison, I have combined the data for both English-derived tags from Table 4.42 above with the figures for the indigenous tags, presented in Table 4.49.

Table 4.48. Pragmatic functions of indigenous tags in the ICE-India conversation files

Pragmatic function	*na*-tag	*no*-tag	Total
Informational	4 (2.13%)	4 (0.94%)	8 (1.31%)
Confirmatory	21 (11.17%)	59 (13.92%)	80 (13.07%)
Facilitating	42 (22.34%)	87 (20.52%)	129 (21.08%)
Attitudinal	102 (54.25%)	252 (59.43%)	354 (57.84%)
Peremptory	4 (2.13%)	4 (0.94%)	8 (1.31%)
Aggressive	1 (0.53%)	–	1 (0.16%)
Signalling entreaty	4 (2.13%)	8 (1.88%)	12 (1.96%)
Unclear	10 (5.32%)	10 (2.36%)	20 (3.27)
Total	188 (100%)	424 (99.99%)	612 (100%)

Table 4.49. Pragmatic functions of all tags in the ICE-India conversation files

Pragmatic function	Canonical tags (total/%)	Invariant tags (total/%)	*no*/*na*-tags (total/%)
Informational	–		8 (1.31%)
Confirmatory	2 (11.11%)	8 (11.76%)	80 (13.07%)
Facilitating	11 (61.11%)	33 (48.52%)	129 (21.08%)
Attitudinal	2 (11.11%)	9 (13.23%)	354 (57.84%)
Peremptory	1 (5.55%)	2 (2.94%)	8 (1.31%)
Aggressive	2 (11.11%)	1 (1.47%)	1 (0.16%)
Signalling entreaty	–	–	12 (1.96%)
Independent utterance	–	13 (19.11%)	–
Unclear	–	2 (2.94%)	20 (3.27)
Total	18	68	612 (100%)

In example (4.6.53), the speakers are trying to find the best date for a party. Speaker A in example (4.6.54) tells speaker B about his father-in-law and his involvement in a court case:

(4.6.53) A: We can have it on Sunday
It's not *Padva* **na** <,>
B: *Padva* is on fourth that is Saturday (ICE-IND: S1A-003#49–51)[97]

(4.6.54) A: My father-in-law is suspended you know that **na** <,,>
B: I don' know <,> (ICE-IND: S1A-041#125–126)

97. *Padva*: the third day of the Hindu *Diwali* festival.

The percentage of tags used for confirmation is slightly higher for *no*/*na*-tags in comparison to the English-derived tags. However, this usage accounts for more than a third of all tokens in Tottie & Hoffmann's BrE data. Some examples from ICE-India follow:

(4.6.55) C: You said you were teaching right somewhere
B. Yeah <,>
C: Which school was that <,>
B: No no it's not computer <,>
A: No but you have taught **na**
C: Taught in some school right
B: Oh yeah <,>
It is uh <,,> Mariam Convent School (ICE-IND: S1A-042#39–46)

(4.6.56) B: No today t[h]ere is a talk **no**
A: But that is <,> in the evening **no**
B: Yeah <,> (ICE-IND: S1A-095#15–17)

(4.6.57) B: *Nahi nahi* <,> you are shifting to which place?
Only Narayan Park **na?**
A: Yeah <,> (ICE-IND: S1A-098#179–181)

The distinction between confirmatory and facilitating tags is not always straightforward. Only one of the three examples above features an interrogative combined with a clause-final *na*/*no*, but the conversational context in which all of the preceding examples are embedded suggests that the speakers were genuinely enquiring for information hitherto unknown to them. Facilitating tags occur in similar discourse contexts, in questions or statements which are directed at other participants in the conversation; but in this case the main function is not epistemic, but affective. The speaker in example (4.6.58), for example, is perfectly aware that there is a holiday – this becomes obvious in the discourse context when she herself supplies the name of the holiday, *Ganesh Chaturthi*. The *no*-tag added to the proposition *tomorrow is holiday* ***no*** is an invitation to the hearer to respond and take a turn in the flow of the conversation.

(4.6.58) A: Okay <,> tomorrow is holiday **no** <,> Ganesh Chaturthi
B: Yeah tomorrow is holiday for us <,> (ICE-IND: S1A-086#17–18)[98]

Most of the time, the speaker who employs a facilitating tag seems to expect and obtain an affirmative answer, as in example (4.6.59) as well as (4.6.60), where there is no unexpected turn in the conversation.

98. *Ganesh Chaturthi:* Hindu religious festival, a celebration of the god Ganesh.

(4.6.59) B: I should really <,> I feel like to go to Kashmir for [one word] <,> Kashmir and Switzerland is same **na** <,>
A: Yeah <,> Kashmir is Switzerland of uh <,> Asia <,,>
It is very beautiful
When we were living <,> when I was living in Kashmir **no** I was brought up there only <,> and everything is
(ICE-IND: S1A-054#252–255)

(4.6.60) A: And you said you were <,> you were in Bombay **no** <,>?
How was Bombay?
B: I did my schooling in Bombay <,,> (ICE-IND: S1A-031#251–253)

However, example (4.6.61) shows that the speaker may also be contradicted: speaker B ventures the perfectly innocent utterance *Today the weather is so pleasant **no***, but does not succeed in eliciting speaker A's positive response.

(4.6.61) B: Today the weather is so pleasant **no** <,>
A: Pleasant <,>
B: Yeah
A: I don't know (ICE-IND: S1A-070#313–316)

Before I turn to the attitudinal tags which account for more than 50 percent of all indigenous tags, I would like to focus on the remaining three categories of tags which are only represented by a handful of examples. Recall that the peremptory and the aggressive use of tags are only represented by 1% of BrE examples in Tottie & Hoffmann's data, and that the available descriptive accounts of Hindi do not include these usages for the *no/na*-tag. The remaining pragmatic category of "signalling entreaty", on the other hand, is derived from the descriptions of Hindi usage and not found in BrE or AmE.

The figures for indigenous IndE peremptory and aggressive tags are even lower than for the corresponding English-derived tags. The two examples (4.6.62) and (4.6.63) may serve to illustrate the peremptory usage. In example (4.6.62), speakers A (a female teacher, 18–25) and C (a male computer instructor, 26–33) are talking about the poor quality of state schools, especially in Maharashtra, where they say the government has been "dumbing down" the syllabus. Speaker C's utterance *Yeah of course that's what I've said **no*** conveys his impatience or unwillingness of having to repeat himself:

(4.6.62) A: Now if you make the system so much easy for the students that they will <,,> in in a way that they are hampering their own uh knowledge <,> as well as not getting down to the English knowledge <,>
Don't you think it is so

C: Yeah it is <,>
A: That <,> we don't even get enough knowledge
C: Yeah of course that's what I've said **no**
(ICE-IND: S1A-042#181–185)

Example (4.6.63) features two tokens of *na*, and it is the second token, following an imperative, which is supposed to convey to the other speaker that the current topic is now off-limits:

(4.6.63) A: I <,> love judie <,>
She's girl only <,> but I feel like boy also <,> while talking <,> with her **na** <,> [laughter]
B: Oh
So you made her uh your boy friend uh
A: Boy and girl friend
B: Oh <,,>
A: Then <,>
B: Then what <,>
A: Enough **na** <,,>
B: Then <,> [laughter]
Then uh <,> then uhm how are your colleagues in office <,>
(ICE-IND: S1A-049#171–181)

The single example of an indigenous tag that might reasonably be classified as "aggressive" runs as follows:

(4.6.64) B: Why don't you introduce me to her **yaar** <,,>
A: Why <,> Why should I
B: Because **yaar** she is good looking let me see if I can <,>
A: Why should I
A: Come on <,> that's why that's why I have not introduced her to you
B: Buddy have faith in me **yaar** <,> I am not such a guy
A: I <,> just now you said **na** <,>
B: Hey <,> it was just for a heck of it <,>
A: Uh okay then <,,> uh sometimes (ICE-IND: S1A-038#134–143)

The two speakers (both male, age 26–33) are initially quarrelling playfully about a woman who is a friend of speaker A. Speaker A is beginning to get exhausted by speaker B's constant references to A's female friend, whom B wants to meet. Thus, with the utterance *I <,> just now you said **na*** he challenges B by reminding him of his perhaps too playful comments about the female friend. Speaker B reacts with a retreat, *Hey <,> it was just for a heck of it*, which speaker A seems to accept. With

exactly one out of 612 examples of indigenous tags (0.16%), aggressive tags are even more marginal among the indigenous tags than among the English-derived tags – and even this tentative generalization should be taken with a grain of salt, since our basis for comparison consists of only four tokens overall.

The pragmatic function of signalling entreaty is realized by 12 tokens, e.g.:

(4.6.65) A: Are you coming to the bazar now <,> with me <,> to get the sweater
B: No <,>
A: No <,> you are not coming <,>
B: Uhm
A: You come **na** <,> please you come <,,>
B: I want to do some library references **na** <,>
(ICE-IND: S1A-059#141–146)

(4.6.66) A: Want to come down <,,>?
B: I'm planning *yaar* <,> let's see <,,>
A: Come **no** Shaukat is here Natalie is here <,> even if Savita is not there they two are there **na** <,> (ICE-IND: S1A-098#266–268)

In both cases, *na* follows an imperative, just as in the Hindi example (4.6.51) above which was given by Y. Kachru to illustrate this particular function of the *no*/*na*-particle in Hindi. Even though this function only accounts for 1.96% of all *no*/*na*-tags in the data, it is highly relevant as evidence that borrowing in this case also included the import of a new pragmatic function for the borrowed form derived from the source language.

Finally, the attitudinal *no*/*na*-tags are much more frequent in the ICE-India data than in Tottie & Hoffmann's data. Attitudinal tags make up 57.84% of all indigenous tags compared with 18% in BrE and 12% in AmE (cf. Table 4.48 above). Admittedly, even if the overall definition of "attitudinal" surely holds for all examples in the category, namely as adding emphasis to the speaker's utterance without an expectation of involvement or reply from the hearer, the label still comprises an array of pragmatic functions that are difficult to pin down for somebody who does not have IndE as his/her native language. It is also in this category where differences in distribution in comparison to the English-derived tags come to the fore. Table 4.50 shows that more than 40% of all *no*/*na*-tokens do not occur in sentence-final position. Valentine's observation on the position and function of *no*/*na* is repeated here (1988: 157):

> These markers used in final utterance position gain a response [...] 100% of the time. When it occurs medially, strictly speaking as a maintenance device, it still evokes a response.

Table 4.50. Syntactic position of indigenous tags in the ICE-India conversation files

	Change of turn	No change of turn	Total
***na*-tag:**			188
in final position	7 (51.59%)	18 (9.57%)	115 (61.17%)
in medial position	47 (25%)	26 (13.83%)	73 (38.82%)
***no*-tag:**			424
in final position	186 (43.86%)	53 (12.5%)	239 (56.36%)
in medial position	124 (29.25%)	61 (14.39%)	185 (43.64%)
Total			612
in final position	283 (46.24%)	71 (11.6%)	354 (57.84%)
in medial position	171 (27.94%)	87 (14.21%)	258 (42.15%)

Table 4.50 further provides the data to quantify Valentine's statement. The tendency that an utterance containing *no*/*na* is followed by a change of turn, that is, evoking a response, is more pronounced for final than for medial *no*/*na* but not excluded for medial *no*/*na*. The IndE particle *no*/*na* is then syntactically more flexible than the canonical tag question, which is restricted to occurring in clause-final position because of its form.

Question tags are clauses and depend on their well-formedness as well as their felicitous occurrence on some preceding clause.[99] Some examples from ICE-GB may serve to illustrate the few highly specific contexts where the question tag is licenced to occur in other positions than strictly sentence-finally or clause-finally. Examples (4.6.67) to (4.6.70) feature tags between clauses:

(4.6.67) Uhm well it's not his fault **is it** if he hasn't uh been sent the forms (ICE-GB: S1A-007#232:1:B)

(4.6.68) Yes the English are branded on their tongue as they say **don't they** so uh as soon as you speak you know they usually know what an idiot you are (ICE-GB: S1A-020#044:1:B)

(4.6.69) It's a funny thought **isn't it** that I was embarrassed (ICE-GB: S1A-032#037:1:A)

(4.6.70) Well then it's your mother tongue then **isn't it** because your mother speaks it (ICE-GB: S1A-069#115:1:B)

Some examples involve right dislocation (cf. Chapter 4.4.2) or vocatives; in these cases, the tag precedes the right-dislocated constituent and the vocative:

99. Examples such as *interesting, isn't it?* do not constitute counterexamples, but are instances of ellipsis.

(4.6.71) I think it actually goes through Brussels **doesn't it** the motorway (ICE-GB: S1A-021#197:2:B)

(4.6.72) Because if you're in London you apply to their operations in London **don't you** <,> you stupid little girl (ICE-GB: S1A-038#246:1:A)

The only examples where the tag question does occur inside the clause are those in an existential construction or in sentences with anticipatory *it*:

(4.6.73) I mean your mother there was a large picture of your mother's mother **wasn't there** in a sort of wig looking as fierce as anything (ICE-GB: S1A-007#167:1:B)

(4.6.74) There is uh There're remarkable stories **aren't there** about these people women <,> (ICE-GB: S1A-063#129:1:A)

(4.6.75) Well I mean it's up to the electorate **isn't it** uhm <,> who gets the posts (ICE-GB: S1A-069#290:2:A)

Similarly, all Hindi grammars that I consulted list *no*/*na* as a clause-final or sentence-final tag. As the examples below will show in more detail, the particle *no*/*na* in spoken IndE may occur clause-internally, following a phrase. I have no way of establishing whether this positional freedom is a feature carried over from Hindi or whether we are dealing with innovative usage that only developed in the host language IndE.

Examples (4.6.76) and (4.6.77) illustrate the most common context for all indigenous IndE tags regardless of their actual pragmatic function. *No*/*na* occurs sentence-finally and elicits some kind of response from the listener, resulting in a change of turn:

(4.6.76) B: In uh <,> we have with <,> we'll have our food there only and come **no**
A: There only <,> (ICE-IND: S1A-051#141–142)

(4.6.77) A: Yeah we had journalism as an ancillary
B. Yeah <,> we too had
B: And I was very interested <,> in journalism **no** <,,>
A: Yeah (ICE-IND: S1A-021#142–144)

However, it is clear in both examples that the speaker is not asking for confirmation from the listener, whose response simply marks his/her ongoing involvement in the conversation.

No/*na* may also occur at a clause boundary in compound or complex clauses, similar to the examples of canonical tags discussed above:

(4.6.78) A: Where did you study <„>?
C: Uhm <,> my father is in military **no** so I studied in <,> uh central school (ICE-IND: S1A-021#25–26)

In (4.6.79), speaker B is a school teacher who has to start teaching again at her school on the fifth, speaker A is stating the obvious but sad fact that speaker B *can't take leave* ***na***. The speakers in example (4.6.80) are reminiscing about their student days:

(4.6.79) B: So other day all of us will leave from Bangalore on fourth <,>
A: Uhm <,>
B: Fourth evening <„>
A: And uh <,> then you can't take leave **na** <,> once you join there
B: Yeah <,> I cannot <„> (ICE-IND: S1A-064#35–39)

(4.6.80) A: And when you were in degree **na** <„> did you read any of Endumari's novels <„>
B: Of course not <,> (ICE-IND: S1A-017#50–51)

Examples (4.6.81) and (4.6.82), like their BrE counterpart (4.6.69) above, illustrate that *no/na* may occur inside an anticipatory *it*-clause or variants thereof, as in (4.6.81) where a nonferential *that* replaces anticipatory *it*. In (4.6.82), speaker A is enquiring about speaker B's recently opened beauty parlour.

(4.6.81) B: But that is also but that is also important **na** how to go about things
If you don't know how to go about things then what you are going to do <„>
A: Uhm that's true (ICE-IND: S1A-003#285–287)

(4.6.82) A: So now how it is running <,>
B: Ah it's okay
A: Ah
B: It also it'll take time **no** to improve <,>
A: That area is quite good no improving
B: Yeah <,> but it will take time to improve <,>
(ICE-IND: S1A-043#45–50)

Some examples occur in those syntactic contexts that have been described in the preceding chapters as prominent in spoken IndE. The first *no*-tag in (4.6.83) as well as *no* in example (4.6.84) occur after the dislocated subject in an LD construction:

(4.6.83) B: Then what is <,> how is are your children <,>
A: My children they are fine <,>
B: What they are doing <„>

A: My <,,> elder daughter Prerna you know **no** <,> she is very intelligent but she is very <,,> lazy
So now-a-days uh I'm telling her you should do all household work also she should learn **no**
B: Yes yes (ICE-IND: S1A-043#68–73)

(4.6.84) B: But prevention is better than cure
A: Yeah that is there but uh <,> most of the people **no** <,> they will eat what they like <,> even it is not good for health
(ICE-IND: S1A-072#88–89)

Example (4.6.85) features a *no*-tag following a subordinate clause as well as another LD construction with *no*:

(4.6.85) And uh really if we read it **no** <,> uh I feel that vegetarian food and uh <,> that *satvik* what you called <,> *satvik* food **no** <,> that will be very good to health (ICE-IND: S1A-072#62:1:A)[100]

Examples where *no*/*na* occur immediately after a constituent below the clause level are also quite common. In examples (4.6.86) and (4.6.87), *no* follows an adverbial.

(4.6.86) A: They won't study <,>
B: Because of the parents attitude **no** we find even the childrens children behave the same way
The students when we try to
A: Yeah they take it lightly (ICE-IND: S1A-085#46–49)

(4.6.87) B: When did you ask <,>?
A: Now only **no** I gave the leave application <,>
(ICE-IND: S1A-086#73–74)

No/*na* may also occur after an NP or PP, as in (4.6.88) and (4.6.89):

(4.6.88) B: At shirdi **na** it's uh uh Shi Saibaba's <,,> shrine
A: Right
Is there any other god other than Saibaba
B: <,> in Shirdi no <,,> (ICE-IND: S1A-012#43–46)[101]

(4.6.89) In uh <,> Mysore **no** they will prefer idli dosa <,> then rice
(ICE-IND: S1A-072#122:1:A)[102]

100. *Satvik* food: a kind of diet associated with Ayurvedic medicine.

101. *Shirdi*: town near Mumbai; *Sai Baba*: important Hindu saint.

102. *Mysore*: city in Karnataka, South India; *idli*: a kind of rice cake, typical of South India; *dosa*: a crisp pancake, equally typical of South India.

Example (4.6.90) again provides a syntactic context that has been discussed in a previous chapter, namely the unlinked topic construction (cf. Chapter 4.3.5). Speaker B takes over from speaker A by repeating her phrase *in the houses* as an unlinked topic, dropping the preposition:

(4.6.90) B: And what do you do on your festivals <,>
A: Yeah I know we are having <,> uh uh different clubs <,> like uh uh town-wise or village-wise and each club is organising dances beach shows and all you know
B: Uhm
Uhm
A: So you can have your choice
Otherwise even on the <,> yeah we are having different beaches and all no people go on the beaches and they enjoy
B: *Accha*
Uh like we have have cer[emony]
A: We don't remain in the houses
B: Yeah yeah
And houses **no** <,> we have <,> pooja and all these kind of things
A: Yeah
Ah we do all that (ICE-IND: S1A-065#161–173)

Finally, speakers may resort to multiple instances of *no*/*na.* The following example (4.6.91) comes from the conversation with the highest number of *no*/*na*-tags (38). Speakers A and B have been exchanging banter throughout the conversation, and when speaker B continues teasing speaker A with *Why you always fall into problem*, A's turn is positively peppered with attitudinal tags:

(4.6.91) B: Why you always fall into problem <,>
A: No I'm do not fall in the problem <,> because I like like this <,> because **na** whenever we are coming from outside **no** <,> then get we fever and all **no** <,> someone should look after **no** then only <,> we can look after others <,> like this I'm doing the work **no** <,> I'll I can't do all the work (ICE-IND: S1A-049#187–188)

Before turning to a further look at the correlation of indigenous tags with speaker variables, it is worth keeping in mind that *no*/*na*-tags are just as much a matter of speakers' personal styles as the English-derived tags. Eight of the ICE-India conversations feature more than 20 *no*/*na*-tags, 205 tokens overall which amount to more than a third of all attested tokens. On the other hand, we find nine conversations with only one *no*/*na*-tag each.

4.6.2.6 *IndE tags: Evidence for fusion. No/na* in spoken IndE is evidently a borrowing and so frequent in comparison to the English-derived tags that we might even be dealing with a case of fusion in Matras' sense (Matras 2000a). Before I turn to a discussion of the relevant notions, I will first take a closer look at the speaker variables that correlate with the distribution of indigenous tags.

All of the parameters "gender", "age", "educational level" and "genetic affiliation of mother tongue" are highly significant for the distribution of *no*/*na*-tags (p = <0,0001). The patterns are similar to the pattern for the invariant tag: younger and female speakers are overrepresented with respect to the occurrence of *no*/*na*-tags.

For those who hold a PhD or a Master's degree, the proportion of indigenous tag is even below the proportion of invariant tags. However, even though there is a marked shift from under- to overrepresentation from the Master's degree category to the Graduate category, the Secondary category does not show such a pronounced overrepresentation.

Interestingly, speakers of Dravidian languages are slightly overrepresented in their use of indigenous tags, just as they were for the invariant tag. This is clear evidence that *no*/*na* has become a pan-Indian form, regardless of its origin from Hindi as the source language.

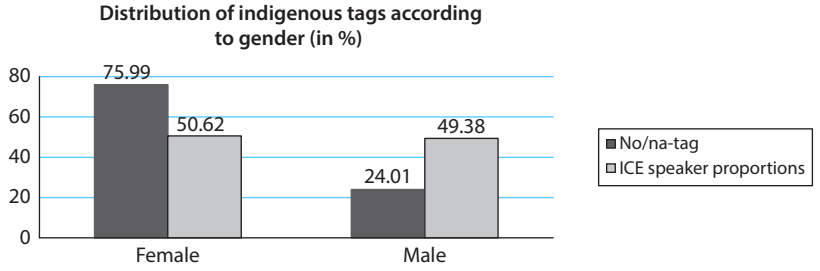

Figure 4.51. Distribution of indigenous tags in the ICE-India conversation files according to speaker's gender

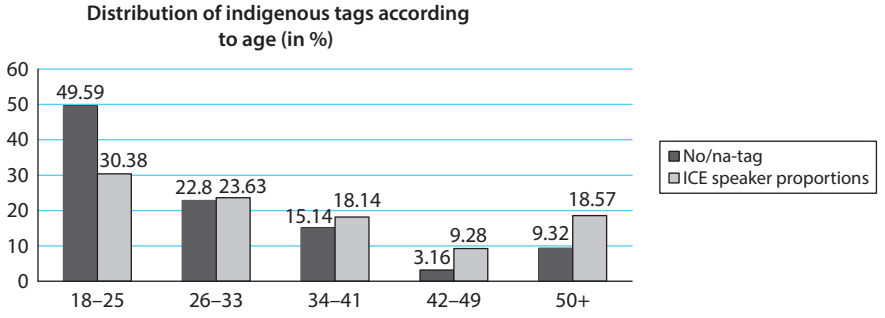

Figure 4.52. Distribution of indigenous tags in the ICE-India conversation files according to speaker's age

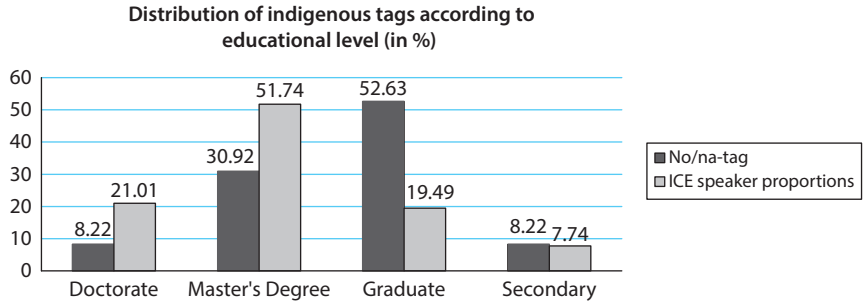

Figure 4.53. Distribution of indigenous tags in the ICE-India conversation files according to speaker's educational level

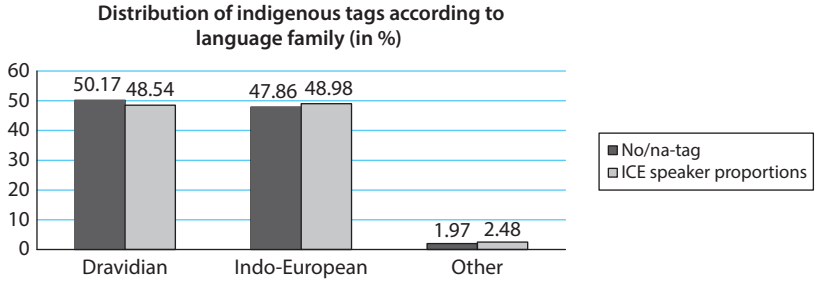

Figure 4.54. Distribution of indigenous tags in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

If we look only at the figures for the *na*-tag which was marked as "indigenous" in the corpus, the proportional preference by speakers of Dravidian languages for the form becomes even more pronounced. Speakers of Hindi are actually underrepresented with respect to *na* as well as *no*/*na* combined. While they comprise 5.81% of the ICE-India speaker population, they only account for 4.32% of the *na*-tags and 4.73% of the overall indigenous tags. *No* as variant of *na* is phonologically more integrated into the recipient language, but both forms must be considered borrowings which belong to speaker's repertoires regardless of their mother tongue. The fact that most Indian languages regardless of genetic affiliation have similar particles functioning as tags has probably facilitated the pan-Indian adaptation of the feature into IndE.

We are dealing, then, with the established borrowing of a bilingual discourse marker in spoken IndE. I have alluded to the ongoing controversy about the distinction between code-switching and borrowing in Chapter 2.3. Our data on English-derived and indigenous tags in spoken IndE provide an interesting opportunity to

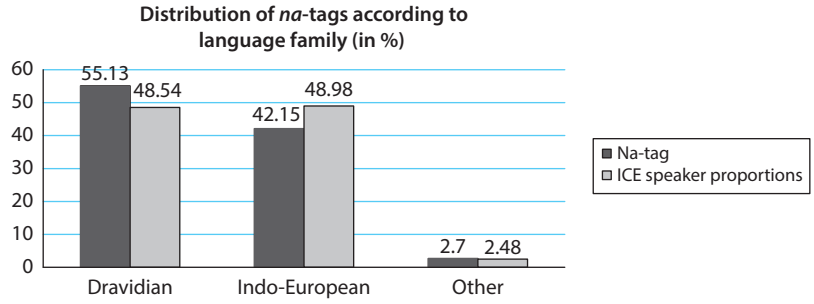

Figure 4.55. Distribution of *na*-tags in the ICE-India conversation files according to genetic affiliation of speaker's mother tongue

revisit the distinction. First of all, the theory of code-switching proposed by Poplack and others (e.g. Poplack 1980, Poplack & Levey 2010) would be misleading here in several respects. Poplack's definition of code-switching (CS) is, rather uncontroversially, as follows (2004: 589):

> CS refers to the utterance-internal juxtaposition, in unintegrated form, of overt linguistic elements from two or more languages, with no necessary change of interlocutor or topic.

Such a definition would force us to make a categorical distinction between *na* as a code-switch and *no* as a phonologically integrated loan where there is in fact no difference in actual usage or in speaker-related preferences. Poplack further distinguishes between two types of code-switching based on the degree of syntactic complexity of the switch. Intra-sentential switches "as a more complex or 'intimate' type" (1980: 589) are contrasted with "emblematic switching" (1980: 590), for example tags, discourse markers, fillers and interjections and idiomatic expressions:

> their insertion in discourse has few, if any, ramifications for the remainder of the sentence. Tags are freely moveable constituents which may be inserted almost anywhere in the sentence without fear of violating any grammatical rule. (1980: 589)

In Poplack's framework, tags and other tag-like switches are indicative of a low degree of bilingual proficiency because they "can be produced in L2 with only minimal knowledge of the grammar of that language" (1980: 605). A preference for emblematic switching including tags is then associated with non-fluent bilinguals on the fringe of the bilingual speech community:

> In-group membership favours intra-sentential code-switching, while non-group membership favours emblematic switching. In other words, that type of switch which all investigators agree to be 'true' instances of code-switching was mainly reserved for communication with another in-group member. (1980: 589f.)

Such less proficient speakers use emblematic switches to "flag" their identity as members of a bilingual community, while the more complex intra-sentential switches may be beyond their bilingual competence. That is, emblematic switches are correlated with a low rather than a high level of bilingualism in a speech community.

Matras has highlighted the diverging approaches to the issue of code-switching versus borrowing when it comes to the study of bilingual discourse markers such as tags (2009: 138):

> It is noteworthy that some authors discuss bilingual discourse markers as codeswitches, attributing to their use special conversational effects such as competence- or identity flagging or the highlighting of discourse boundaries, while others treat them as borrowings or indications of ongoing language change. It appears that we are dealing with an area in which frequent insertional switching has the potential and the tendency to become stabilised and over time to lead to language change.

The emblematic function of flagging that Poplack and others associate with tag switches would have, in effect, a metalinguistic function. The speaker's tag switches are seen as becoming meaningful not primarily in the immediate discourse context, but on a more general interpersonal level when the speaker signals his/her allegiance to a specific speech community. Matras casts some doubts upon the concept of emblematic switches and the notion of flagging, since there is simply no need for flagging allegiances in a multilingual environment (2009: 143):

> As for the need to flag bilingual competence, it fails to explain the adoption of foreign discourse markers in situations where bilingualism is taken for granted and there is little to be gained by flagging acculturation.

However, considering *no*/*na* as emblematic switches also has its merits in the Indian context. After all, d'Souza's characterization of *no* as "ubiquitous" (1991: 312) quoted in the beginning of this chapter is derived from her study of speech acts in IndE fiction, where the frequent use of *no* serves to evoke a specific "Indianness" for Indian authors writing in English. The social meaning that is carried by emblematic switches in writing is, however, different from that in direct conversation. In literary language, flagging is not related to the pragmatics of speakers' interaction, but to the text itself. The text as a whole is situated within a specific cultural context, i.e. flagging is intertextual rather than interpersonal.

As already mentioned, the more interesting question in this area is why discourse markers are so readily borrowed in language contact scenarios. It is uncontroversial that *no*/*na* must have entered IndE as code-switches. The astonishing fact is rather the sheer frequency of the forms in a corpus of standard spoken IndE despite their reputation as "bad English". Moreover, with only a handful of tokens,

the canonical tag seems to be on its way out of the repertoire of spoken IndE. The invariant tag as the other English-derived tag occurs more frequently, but still remains a minority option with regard to the indigenous tags.

We may therefore tentatively assume that we are witnessing a case of fusion in spoken IndE, defined as "the wholesale non-separation of languages in both forms and functions of a given category or class of expressions" (Matras 2000a: 84). As noted at the beginning of this chapter, tags and other utterance modifiers are particularly susceptible to borrowing because of their special status and function in a "grammar of directing" (Matras 1998: 325):

> discourse operators or utterance modifiers carry out highly automated routine tasks, for which routine schemas appear to exist. Their role in managing the interaction and in mentally reaching out to the hearer gives them a gesture-like function. As such, they are less identifiable with just a particular subset of the repertoire and its permissible usage contexts – i.e. less attributable to a particular 'language' – and are instead accepted as more universal. They are, in other words, 'pragmatically detachable' from their source language. (Matras 2009: 140)

The reason for the dominance of indigenous over English-derived tags in spoken IndE is then neither speakers' lack of proficiency in English nor a need to flag some cultural identity, but related to the special cognitive status of utterance modifiers in the grammar of multilingual speakers. I will return to an assessment of the relevance of this observation for the overall grammar of spoken IndE in the final chapter.

CHAPTER 5

Conclusion

The preceding chapter has dealt in some detail with a specific domain of IndE syntax on the basis of a specific text category. I investigated a cluster of features related to discourse organization in the ICE-India private conversation files, i.e. in a corpus of standard conversational IndE. My justification of these choices is derived from a set of intertwined methodological premises which in turn informed my research questions. In this chapter, I will first recapitulate the theoretical underpinnings of my corpus-based study and then summarize my findings concerning the syntactic profile of spoken IndE. In a next step, I will evaluate my results by situating them in the broader contexts of research on New Englishes and on contact-induced language change. One question cuts across both these contexts: which of the features described in the preceding chapter has the potential to diffuse to a wider range of text types and thus to become part of an emerging standard Indian English? My tentative answer will again be largely theoretical, as this study makes no pretences whatsoever to being relevant for applied linguistics or language pedagogy. However, I do believe that my study has significantly advanced our understanding of the norms of spoken Indian English. It is for the IndE speech community to decide if and how these descriptive norms are extended to become prescriptive norms and thus (part of) the standard.

5.1 Indian English as a contact language

Mesthrie has repeatedly called for "a greater degree of rapprochement between the fields of world Englishes (or New Englishes) and Contact Linguistics" (2006b: 273). In an earlier article in the same vein, he added that "attention to certain aspects within the variationist enterprise in sociolinguistics [...] can only strengthen both areas" (2003: 450). As Mesthrie stresses, in order to establish "a truly comparative data base for linguistic analysis" (2006b: 273) in the field of New Englishes, researchers should turn their attention to spoken data elicited in informal contexts in line with established sociolinguistic methodology and "not mix spoken and written data" (2006b: 274). This suggestion is probably quite uncontroversial, at least from a sociolinguistic perspective, but by no means trivial. In the World Englishes paradigm as suggested by Mesthrie, the familiar

sociolinguistic assumption that the vernacular is the 'real thing' receives another dimension from the field of contact linguistics. It is in actual multilingual conversational encounters that we can observe the impact of language contact on speaker's linguistic repertoires, and it is these communicative situations that may trigger contact-induced language change. This perspective on language contact has been advocated by Matras, whose theoretical framework has proven invaluable for the present study (cf. Chapter 2.3). Matras distinguishes the actuation of innovative features in multilingual communication strictly from the propagation of such contact-induced language change. The former is a cognitive, the latter a social process which is subject to quite different constraints than the former. To be sure, instances of actuation are not amenable to linguistic analysis: contact-induced features will only become apparent as such once they "have become favoured by a significant number of speakers in a significant set of communicative constellations" (Matras 2007a: 52).

Still, for anybody who wishes to investigate spoken IndE along these lines, ICE-India is the obvious choice. The corpus as a whole is the only publicly available balanced corpus of IndE, and the conversation files S1A 1–100, on which my study is based, have the added advantage of including a broad range of speakers of IndE. Since information about age, gender, mother tongue and educational level is included for each speaker, it becomes possible to draw on the methods of variationist sociolinguistics and thus to meet Mesthrie's demand for a more integrated approach in the study of World Englishes. Other IndE corpora that have been created in recent years, while also belonging to the individual authors, have been compiled with different objectives in mind. Sedlatschek's corpus of IndE includes 80,000 words of broadcast transcripts as the only spoken text type (2009: 43). Balasubramanian's corpus of contemporary IndE (CCIE) contains a category of 'Conversational English', which amounts to 233,912 words. This category includes 'Oral Interviews', 'Service Encounters', 'Spoken Entertainment' and only nine actual conversations (65,324 words) in which "the majority of the participants were South Indian" (2009b: 55).[1] Thus, even though the overall ICE-project was well under way before the 'contact-linguistic turn' in the study of World Englishes gained momentum, the individual ICE-corpora represent a so far unsurpassed database for the investigation of spoken English in multilingual contexts. One important aspect of my study concerns the basis of comparison for judgments of frequency of a particular construction. I could draw on the parallel sections of ICE-GB to analyze the frequencies and discourse contexts of the constructions under scrutiny. By doing so, I was first of all able to give intuitive judgments about

1. Interestingly, the only corpora of *spoken* IndE were compiled in expatriate communities, cf. Sharma (2003) for IndE spoken by immigrants to California and Mesthrie (1992) for SAIE.

the higher frequency of specific constructions in IndE a sound empirical foundation. I was further able to avoid a conceptual trap that has loomed large in the field (cf. Chapter 2.2.2). Rather than comparing spoken IndE usage with an idealized BrE standard as laid out in e.g. the *Comprehensive Grammar of the English Language* (Quirk et al. 1985), I compared IndE performance data with parallel data from BrE. Again, this would not have been possible without the ICE-project.

The linguistic features that I examined in spoken IndE all contribute to the domain of discourse organization or discourse management. All of them can be employed by speakers to express discourse-pragmatic meaning. Speakers may manipulate canonical word order to foreground or background salient information according to their point of view, they may mark sentence focus morphologically by adding a focus particle, or they may use discourse markers to express a range of discourse-pragmatic meanings. Two aspects were important for delineating the topic of investigation in this way. First of all, the features under discussion have previously been described as either 'typical' of New Englishes in general or as unique to IndE. The features that are supposedly common to the New Englishes include higher frequencies for topicalization and left dislocation constructions as well as the occurrence of invariant tags. The two features that are not attested in other New Englishes are the non-initial existential *there*-construction and *only* as presentational focus marker.[2]

Taken together, the features point to one of the main challenges in the field of New Englishes: how do we account for the similarities in innovations across New Englishes, given the widely varying local contact languages? When a linguistic feature occurs e.g. in an Asian English as well as in an African English, substrate influence as the main determinant becomes unlikely, given the sheer number of indigenous languages across Asia and Africa, not to mention their typological diversity: "It is prima facie implausible, areal linguistics notwithstanding, that over a thousand languages should induce the very same (or similar) influences" (Mesthrie 2008: 634). The default explanation in such cases has then been either "universals of second language acquisition" or reference to "Angloversals". Similarly, if a feature is only present in one variety, is it then necessarily the product of transfer?

Recently, important steps towards a solution to this puzzle have been taken. One solution has already been mentioned repeatedly, namely the reappraisal of the notions "substrate influence", "interference", "transfer", or "borrowing" (cf. also Chapter 2.3.2). The perspective on 'interference' that used to be dominant in the field was derived from second language acquisition research, where it was associated with imperfect language learning (cf. Chapter 2.2.2). Over the last years, this

2. Presentational *only* does occur in SAIE, another 'New Indian English', to be precise (cf. Chapter 4.6.1).

perspective has shifted. Multilingual speakers are no longer seen as "slaves" of their mother tongue, passively and indiscriminately mapping structures from their first language onto a target language. Consequently, transfer or borrowing are no longer taken to indicate that a speaker has not mastered the target language fully, but rather as "a common creative resource available to ease the tension between interacting linguistic systems, whether related or unrelated" (Johanson 2008: 62). This process of "creative restructuring [of the target language] according to cognitive principles and communicative needs" (2008: 63) may lead to similar results in multilingual contact scenarios, even when the languages sharing the relevant communicative spaces are typologically very different.

Another call for a more careful consideration of the apparent similarities across New Englishes has come from the contributors to the special issue of *English World-Wide* on "The typology of Asian Englishes" (cf. Lim & Gisborne 2009). All authors strongly suggest a more refined approach when arguing for or against universals vs. substrate influence in the New Englishes: "Surface similarities across New Englishes can be skin deep, diverging dramatically upon closer examination, due to substrate systems or substrate-superstrate interaction" (Sharma 2009: 190).

In my study, I combined these two approaches, both on the conceptual and the empirical level. Living in a multilingual society like India provides evidence for the creative crossover between linguistic repertoires in communicative interaction practically by the minute. I have shown in Chapter 3.2 that "linguistic repertoire" should not be too narrowly defined in a country where grassroots multilingualism is pervasive. Conceptually, this means that there is no need to isolate exactly one language within the Indian communicative space and then argue for substrate influence deriving from that language. Concomitantly, this also means that substrate influence is not *per se* ruled out just because the Indian communicative space comprises an impressively large number of different languages. In a sociolinguistic area such as India or South Asia generally, the notion of 'substrate influence' further has to be extended beyond the classical preoccupations of areal typology, such as diffusion of phonological, morphological and syntactic features. The next chapter will describe in more detail the impact of *pragmatic* sprachbund features on IndE.

My second reason for investigating the cluster of features expressing discourse organization is related to these considerations. I assumed that speakers in multilingual scenarios will employ similar communicative strategies within a unified domain. For example, if speakers who want to change the canonical topic-comment structure of a proposition do so by replicating a pattern from their source language, their doing so will probably not be restricted to just one particular construction belonging to the domain of discourse-pragmatic sentence structure. In

this I slightly diverge from Thomason's agenda for establishing contact-induced language change:

> First of all, in order to establish that contact-induced change has occurred, it is necessary to look at the languages as wholes: if structural interference of any kind has occurred, it is highly unlikely to be isolated in the system. The best strategy is to look for a series of independent, unrelated structural changes in the proposed receiving language. (2008: 49)

Even though I do agree with Thomason that a claim for contact-induced language change should not be made on the basis of one single structural feature in a contact language, I suggest we should look for *related* rather than *unrelated* structural changes. The adoption of a specific contact-induced strategy within a specific domain is likely to have repercussions within the domain as a whole, affecting similar strategies and possibly leading to a restructuring of the whole domain. So far, my hypothesis has been confirmed, as will become apparent in the next section.

5.2 Multilingual competence: The norms of spoken Indian English

Three of the linguistic features belonging to the domain of discourse organization involve a reordering or realignment of basic sentence patterns (cf. Chapter 4.1.5): the non-initial existential *there*-construction, topicalization and left dislocation (LD). The latter two are regular features of all varieties of English, albeit at a considerably lower frequency, while the non-initial existential construction is an innovation in spoken IndE. However, considering these three constructions as part of a unified domain involves more than just registering the higher frequency of topicalization and LD constructions. After all, the overall perspective on language contact adopted here takes for granted that there "are always dissimilarities in substance, meaning, contextual applicability and *frequency* between models and copies" (Johanson 2008: 62f., emphasis mine).

I analyzed the form, function and discourse contexts of all three constructions in detail in order to arrive at speakers' motivations and strategies for employing a particular linguistic structure in context. One overarching discourse motivation for all three constructions that emerged from the analysis is *topic continuity*. That is, "the pragmatics of information structure has lost its significance" (Heine 2008: 51) for a considerable proportion of the constructions analyzed. Rather than expressing some diversion from the canonical topic-comment structure, many examples served the innovative function of creating textual cohesion by repetition of the immediately preceding topic NP. This strategy has been labelled "elliptical repetition", a pan-South Asian politeness device in conversation (cf. Chapters 4.2.6.1, 4.3.6 and

4.4.5). Thus, the South Asian "grammar of culture" (cf. Chapter 3.2.4) provides the impetus for a re-interpretation and eventual unmarking of three pragmatically marked word order patterns. Such a "contact-induced word order change without word order change" (cf. Heine 2008) is not uncommon across languages, as Heine and others have duly noted. Heine further points out that this process is typically situated at the syntax-pragmatics interface:

> even if there is a clearly syntactic goal, such as copying a word order characteristic of another language, *the strategy employed to achieve this goal is not really syntactic but rather semantic or pragmatic in nature*. (Heine 2008: 57, emphasis mine)

The non-initial existential *there*-construction is interesting because it represents the only innovative IndE structure among the three patterns, while sharing with them the innovative discourse function of indicating topic continuity. I have shown that the construction is a pattern replicated on the model of the Indian languages while still preserving the structural integrity of the target language English. Thus, language contact is manifest at the level of discourse functions in all three constructions, while the non-initial existential additionally provides evidence for language contact at the level of form.

For topicalization, left dislocation and non-initial existential constructions, pattern replication is in effect the only option for speakers to import new discourse functions into IndE (cf. Matras 2009: 237). The remaining features in the domain of discourse organization are not predisposed towards pattern replication by virtue of their structure. The IndE focus marker *only* represents an interesting intermediate case. On the one hand, the form patterns with cleft constructions in being a focus rather than a topic construction like the three preceding constructions. On the other hand, *only* is a morphological marker structuring a proposition and thus overlaps in function with some uses of the invariant tags.

I have already noted that the most remarkable fact about *it*-clefts in spoken IndE is their virtual absence, and I hypothesized that speakers of IndE prefer to mark the focus of a proposition morphologically rather than syntactically or by intonation. *Only* is a calque of the enclitic focus markers common across Indian languages. In the process of calquing, the polysemy of the source was imported into IndE, and *only* now serves the double function of marking both contrastive and presentational focus in IndE. The sub-domain of focus marking within the overall domain of discourse organization, then, has undergone restructuring in IndE: unlike in BrE, morphological focus marking can be applied to two different kinds of focus, and *only* now serves functions that are otherwise expressed by syntactic means or by intonation.

Two questions may be raised in this respect. First, why was *only* calqued rather than borrowed, and second, why was a focus marker calqued and not the

corresponding topic marker, which is also very common as an enclitic focus particle in Indian languages? The first question is easier to answer than the second. Even acknowledging that there are no absolute constraints on borrowing in language contact, researchers still agree that bound morphology is generally highly resistant to borrowing (cf. Matras 2009: 155f.), which leaves the option of calquing to carry over the functions of the enclitic focus marker into IndE. The motivation to do so, in turn, arises from the discourse functions of *only*. Again in Matras' terms, the focus marker *only* belongs to "the class of monitoring-and-directing operators, in that they process hearer-sided expectations and presuppositions" (2009: 99). Within this class, items that express contrast such as *only* are particularly prone to transfer:

> We may hypothesize that this has to do with the conversational tension surrounding the use of expressions of contrast, change and restriction, leading to a high processing overload which is then compensated for through the elimination of the choice between competing expressions in a speaker's multilingual repertoire. (Matras 2000b: 570)

Topic markers would then not be as highly susceptible to borrowing as focus markers. A further factor which might impede the calquing of topic markers is their highly abstract structural meaning. We have seen in Chapter 4.6.1.4 that most Indian languages have a three-way paradigm of enclitic topic and focus markers. The additive focus marker is generally rendered as *also* and the restrictive focus marker as *only*, whereas no such transparent translation is available for a topic marker. Speakers of IndE then lack both the discourse motivation as well as a matching pivot in the target language to create a new morphological topic marker in IndE.

Finally, I have shown that the invariant *no*/*na*-tag in spoken IndE may even be seen as a case of fusion, in which a bilingual discourse marker is not only borrowed, but on its way to replace its original equivalent in the target language. *No*/*na* assumes the same functions in IndE as the canonical tag, and there are some examples of an additional function that has been imported from the source languages in the process of matter replication. It is important to note here that the main motivation for fusion in the domain of discourse markers is again cognitive and neither social nor related to simplification strategies associated with the process of second language learning. I could show that the common characterization of bilingual discourse markers as 'emblematic switches' which are used to 'flag' in-group membership does not make sense in speech communities where multilingualism is taken for granted. That is, there might still be a social reason for preferring *no*/*na* over English-derived tags, but it is of secondary importance. Similarly, the use of invariant tags across New Englishes has often been accounted for in terms of language

learners' generalization and simplification strategies (cf. Williams 1987). Again, the main issue with respect to invariant tags in IndE is not the preference for invariant *isn't it*, but the dominance of *no*/*na* as a bilingual discourse marker.

The preceding discussion has shown that spoken IndE, at least within the domain of discourse organization, bears all the hallmarks of a vibrant contact language. Moreover, the most persistent correlation with speaker variables concerned gender and age: most features under discussion were more frequent with female and younger speakers. Any sociolinguist's suspicion would then be that we are dealing with a case of language change from below, in line with Labov's "*Principle II: In change from below, women are most often the innovators*" (1990: 215, italics in the original). This principle together with Labov's Principle I ("In stable sociolinguistic stratification, men use a higher frequency of nonstandard forms than women" (*ibid*. 205)) has been confirmed time and again as an explanatory parameter in accounts of language variation and change (cf. Labov 2001). Even though Labov's generalizations as well as the bulk of research in the variationist paradigm are concerned with phonological variables, studies such as Cheshire (2005) and Tagliamonte & D'Arcy (2009) have conclusively shown that syntactic and even discourse-pragmatic variables are by no means beyond a variationist approach. A real-time study based on a new corpus of 21st century IndE could then settle the question whether the younger and female speakers in the ICE-India corpus are indeed the agents of language change from below.

The speakers' mother tongue and educational level, while often statistically significant in a very general sense, do not influence the patterns of occurrence as much as many researchers – and many Indian speakers of English, for that matter – might have expected. As indicated in Chapter 1, linguists as well as the speakers of IndE themselves frequently call the existence of a supraregional IndE into question, referring instead to Marathi English, Punjabi English, Telugu English etc. Kachru's familiar "parameters determining variation" in IndE or South Asian English generally come to mind (cf. also Chapter 3.3.3):

> The first is the users' proficiency in English in terms of language acquisition and years of instruction in the language. The second is the region of South Asia to which the user belongs and the impact of the dominant language of that region on English. (Kachru 1994: 508)

Kachru's first parameter of variation is also frequently conceptualized as the "cline of bilingualism", with proficient speakers/users of IndE approximating standard BrE in everything but pronunciation, and learners hovering in between the basilect and the mesolect. This is surely an accurate and useful characterization of the levels of English usage across India. However, the two parameters taken together imply that proficient speakers/users speak a standard English that is largely free

from contact features – that speakers have 'overcome' the influence of their mother tongues, so to say. I was able to show that for the domain of discourse organization in conversational IndE, there is abundant evidence for contact-induced usage patterns, and it is precisely the use of these contact features rather than their absence which contributes to the norms of standard spoken IndE. To repeat, there is a pan-Indian syntax of discourse organization rather than a Hindi syntax, a Tamil syntax, or a Manipuri syntax of discourse organization. Further, individual linguistic features belonging to the domain are not persistently stratified according to educational level, which goes against the common assumption that language contact features are indicative of limited proficiency in English.

Another important point has to be made in this respect: it is simply not true that the main differences between the "New" and the "Old" Englishes are quantitative rather than qualitative. Such an impression comes from the failure to acknowledge the importance of discourse traditions and their specific configurations of orality and literacy, of universal structural parameters of spoken and written language use. If we take a naïve corpuslinguistic perspective on a specific variety, we might well be content to state that there are overall only a few quantitative differences in the occurrence of some features or patterns, with spoken language displaying more variation than the written registers. Such a view ignores the fundamental fact that language contact is actuated in multilingual encounters of actual speakers. Other text types that are further removed from the immediate communicative situation will, in their turn, show progressively more influence of the exonormative standard, which, like all standards, reduces or blocks variation.

My analysis has shown that conversational IndE displays a range of features that are innovative and qualitatively different from other varieties of English. I was only able to do so by combining a quantitative with a context-sensitive approach, by going beyond the sheer frequency of a feature to analysing its function in conversational interaction. These functions point to shared norms of the IndE speech community in the domain of discourse organization, which are ultimately the shared norms of the Indian speech community as a whole, replicated in IndE. Whether these norms will have an impact beyond their immediate discourse context is a different matter, to be discussed in the next section.

5.3 From norms to standards: Indian English in the 21st century

My study was devoted to an analysis of specific features of spoken IndE drawn from the domain of discourse organization. What, then, do we now know about standard IndE? Put differently, which of the features discussed in the preceding section are likely to be selected for an emerging standard?

By asking these questions, I might appear to be moving on very shaky grounds indeed. For one thing, trying to make predictions about future linguistic developments is naturally open to the charge of subscribing to teleology, an unforgivable scholarly misdemeanour. Next, it is for the IndE speech community to decide if there will be a standard IndE, and what form(s) is should take. Making predictions concerning the standard from an outsider's perspective might then appear as another kind of unwarranted exonormative interference. It is therefore necessary to be very precise about the theoretical and methodological assumptions which inform the following discussion.

To begin with the uncontroversial: the ICE-India speaker sample which supplied the data for my study represents educated IndE – even over-educated IndE, considering that more than 90% of all speakers have a university degree (cf. Chapter 4.1.3.5). However, Sridhar, among others, has raised doubts about the simple equation "educated IndE = standard IndE". His reflections about a future standard IndE were made in the context of his study about a proficient speaker of IndE whose mother tongue is Telugu, a Dravidian language:

> no matter how Standard Indian English is eventually defined, it is not enough to say that it is the English used by educated Indian speakers of English. [...] Maybe, by Standard Indian English is meant a variety that is not so regionally marked, and yet it would still carry certain characteristics that would be 'pan-Indian' (and there are several such features, in syntax as in phonology and other levels, due to linguistic convergence [...]) and that would serve to separate it from other varieties of native and non-native English. (1992: 147)

Sridhar was reluctant to label this particular speaker's variety of IndE "standard IndE" because of its "regionally marked" syntactic features. Note that Sridhar explicitly includes 'nativized' features in his concept of standard IndE, but such nativized features on different linguistic levels should be supraregional, or pan-Indian. This requirement is clearly met by the linguistic features targeted in my study. All features under consideration belong to a pan-Indian IndE repertoire, where differences in distribution with respect to speakers' age, gender, education, or mother tongue are more of an indication of the onset of Schneider's stage 5 – dialect differentiation – than an indicator of limited distribution within the IndE variational space. Such a supraregional distribution of features is of course a necessary, but not a sufficient condition for eventually being selected as an element of standard IndE.

I have argued in Chapter 2.4 above that the logic of standardization implies that standard languages are written languages. Put differently: if there will be an endonormative standard IndE, then it has to be a written and not only a spoken standard. Even if the norms of spoken standard IndE as exemplified in my study

for a particular domain become coextensive with standard spoken IndE, there is still the largely exonormative BrE standard for the written mode. The question posed at the beginning of this section can now be made more precise: which of the features I identified as belonging to an emerging standard for spoken IndE also have the potential to become part of an emerging standard for *written* IndE?

Again, I have to be more precise when considering each individual feature's potential for membership in the standard. As said at the beginning of the preceding section, the two features topicalization and left dislocation are found in all varieties of English, but at a lower frequency. For these features, we have to focus not on their form, but on their innovative functions and discourse contexts. The non-initial existential construction is innovative in its form, but just extends rather than violates the general English surface syntax. The case of the utterance modifiers *only*, *itself* and *na*/*no* is different, as these replicated forms are much more transparent to speakers as elements of a source language. Typically, speakers are more conscious of replicated matter than of replicated structural patterns, so that the former are much more likely to be noticed and then openly negotiated within the speech community when it comes to admission to the standard.

Still, all features, albeit to different degrees, conform to the constraint on preserving the integrity of the target language English (cf. Matras & Sakel 2007 and Chapter 2.3.2). Even the *na*-particle, obviously a borrowing, is on its way to being morphologically integrated into the target language as *no*, homophonous with the English negative. The presentational focus markers *only* and *itself* import an entirely new category into English, but do so via an extension of the patterns available for focus marking generally. That is, there is a kind of a cline of inconspicuousness for the features under discussion, with the syntactic features being least obvious and the morphological features more obvious to the IndE speakers themselves as divergences from a BrE standard. I assume that speakers will find it more 'natural' to carry over a specific feature into writing that is not transparently a contact feature unless they want to create a specific effect, for example in creative writing.

However, the one discourse function which cuts across all three syntactic features under discussion is establishing topic continuity by elliptical repetition. Repeating part of the preceding speaker's proposition is a pan-Indian politeness strategy which creates cohesion in interaction. Such a strategy together with the syntactic means to realize it is unlikely to find its way into the written mode, which has altogether other means to establish cohesion that are not primarily geared at cohesion on the interpersonal level. My prediction is, then, that topicalization, left dislocation and the non-initial existential will remain within the realm of the spoken standard. It is possible that the progressive unmarking of pragmatically marked word order patterns will lead to greater variability in the range of basic word order

patterns, but then only when it comes to the position of non-arguments. The reason for this is again the presence of the written standard, as I will show.

It has long been recognized that word order is particularly susceptible to language contact. Thomason's well-known borrowing scale predicts that "syntactic features (such as word order) are the next easiest thing to borrow" (2001: 69), following the borrowing of what she calls "nonbasic vocabulary". Clause-level syntax is more prone to contact-induced language change than phrase-level syntax precisely because of its function of expressing information structure via reordering or realignment of arguments. If a non-canonical word order is used with increasing frequency, it gradually loses its special discourse-pragmatic function and becomes unmarked, eventually even replacing the older basic word order pattern. Such a change in basic word order is, however, only feasible in the absence of a codified written standard. Still, the greater variability in the position of adjuncts that I have noted in passing in several chapters may find its way into written IndE, as there are no clear-cut rules for the position of adverbs, only tendencies and collocational preferences across varieties of English.

The discourse functions of the utterance modifier *no*/*na* are also restricted to spoken language, more precisely to the management of conversational interaction, expressing epistemic and affective meanings. These functions – just like those of its counterpart, the canonical tag – are again unlikely to find expression in the written mode.

The focus markers *only* and *itself* represent an intriguing case, pointing to another set of constraints on the standardization of contact-induced innovations. So far, I have argued that the features under discussion will not become part of a standard *written* IndE because they firmly belong to the repertoire of *spoken* IndE, or of conceptually oral language in general. I have already shown elsewhere that presentational *itself* is a rare form, but does occur in written IndE, while presentational *only* is mostly restricted to direct conversation (cf. Lange 2007). What, then, does *itself* have that *only* doesn't? I would hypothesize that *only* is more transparent as a calque to speakers of IndE than *itself*, which is more abstract in meaning and therefore has less potential than *only* to be targeted by language attitudes which frown upon borrowings. The reason for selecting *itself* rather than *only* would then be social rather than structural.

The future development of IndE will undoubtedly see many more such selection processes. Meanwhile, I look forward to doing this study all over again in twenty years' time – on a new corpus of 21st century IndE.

References

Agnihotri, R.K. 1999. India: English in India. In *The Yearbook of South Asian Languages and Linguistics*, R. Singh (ed.), 184–195. New Delhi: Sage.

Agnihotri, R.K. 2002. Sociolinguistic theory and practice: The Indian counterpoint. In *The Yearbook of South Asian Languages and Linguistics*, R. Singh (ed.), 11–25. New Delhi: Sage.

Agnihotri, R.K. 2007a. *Hindi: An Essential Grammar*. London: Routledge.

Agnihotri, R.K. 2007b. Identity and multilinguality: The case of India. *In Language Policy, Culture, and Identity in Asian Contexts*, A.B.M Tsui & J.W. Tollefson (eds), 185–204. Mahwah, New Jersey: Lawrence Erlbaum.

Algeo, J. 1990. It's a myth, innit? Politeness and the English tag question. In *The State of the Language*, C. Ricks & L. Michaels (eds), 443–450. Berkeley CA: University of California Press.

Algeo, J. 1988. The tag question in British English: It's different, i'n'it. *English World-Wide* 9: 171–91.

Anand, S. 1999. Sanskrit, English and Dalits. *Economic and Political Weekly*, July 24–30 1999. <http://www.ambedkar.org/News/hl/SanskritEnglish.htm> (17 May 2011).

Anderson, B. 2006. *Imagined Communities. Reflections on the Origin and Spread of Nationalism*, 2nd rev. edn. London: Verso.

Annamalai, E. 2004. Nativization of English in India and its effect on multilingualism. *Journal of Language and Politics* 3(1): 151–162.

Ansaldo, U. 2009. The Asian typology of English: Theoretical and methodological considerations. *English World-Wide* 30(2): 133–148.

Asher, R.E. 2008. Language in historical context. In *Language in South Asia*, B.B. Kachru, Y. Kachru & S.N. Sridhar (eds), 31–48. Cambridge: CUP.

Asher, R.E. & Kumari, T.C. 1997. *Malayalam*. London: Routledge.

Austin, G. 1966. *The Indian Constitution: Cornerstone of a Nation*. Oxford: Clarendon Press.

Backus, A. 2004. Convergence as a mechanism of language change. *Bilingualism*: *Language and Cognition* 7(2): 179–181.

Backus, A. 2005. Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism* 3–4: 307–340.

Bauer, L. 2002. *An introduction to international varieties of English*. Edinburgh: EUP.

Balasubramanian, C. 2009. *Register Variation in Indian English*. Amsterdam: John Benjamins.

Bhatia, T.K. 1987. *A History of the Hindi Grammatical Tradition: Hindi-Hindustani Grammar, Grammarians, History and Problems*. Leiden: E. J. Brill.

Bhatt, R.M. 1995. Prescriptivism, creativity and World Englishes. *World Englishes* 14(2): 247–259.

Bhatt, R.M. 2000. Optimal expressions in Indian English. *English Language and Linguistics* 4: 69–95.

Bhatt, R.M. 2002. Experts, dialects, and discourse. *International Journal of Applied Linguistics* 12(1): 74–109.

Bhatt, R.M. 2007. On the native/non-native distinction. In *Annual Review of South Asian Languages and Linguistics 2007*, R. Singh (ed.), 55–71. Berlin: Mouton de Gruyter.

Bhatt, R.M. 2008. Indian English: Syntax. In *Varieties of English 4: Africa, South and Southeast Asia*, R. Mesthrie (ed.), 546–562. Berlin: Mouton de Gruyter.

Bhatt, R.M. & Mahboob, A. 2008. Minority languages and their status. In *Language in South Asia*, B.B. Kachru, Y. Kachru & S.N. Sridhar (eds), 132–152. Cambridge: CUP.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Birner, B.J. & Ward, G. 1998. *Information Status and Noncanonical Word Order in English* [Studies in Language Companion Series 40]. Amsterdam: John Benjamins.

Breivik, L.E. 2005. The historical development of existential *There*: An example of subjectification in grammaticalisation. In *Contexts-Historical, Social, Linguistic*, K. McCafferty, T. Bull & K. Killie (eds), 163–185. Bern: Peter Lang.

Breton, R.J.-L. 1997. *Atlas of the Languages and Ethnic Communities of South Asia*. New Delhi: Sage.

Brinton, L.J. 1998. "The flowers are lovely; only, they have no scent": The evolution of a pragmatic marker. In *Anglistentag 1997 Giessen: Proceedings*, R. Borgmeier, H. Grabes & A.H. Jucker (eds), 9–33. Trier: Wissenschaftlicher Verlag.

Brinton, L.J. 2006. Pathways in the development of pragmatic markers in English. In *The Handbook of the History of English*, A. van Kemenade & B. Los (eds), 307–34. Oxford: Blackwell.

Bruthiaux, P. 2003. Squaring the circles: Issues in modeling English worldwide. *International Journal of Applied Linguistics* 13(2): 159–178.

Calude, A.S. 2008. Clefting and extraposition in English. *ICAME Journal* 32: 7–33.

Chand, V. 2010. Postvocalic (r) in Urban Indian English. *English World-Wide* 31(1): 1–39.

Cheshire, J. 2005. Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers. *Journal of Sociolinguistics* 9(4): 479–508.

Coelho, G.M. 1997. Anglo-Indian English: A nativized variety of Indian English. *Language in Society* 26: 561–589.

Cohn, B.S. 1986. *Colonialism and Its Forms of Knowledge: The British in India*. Princeton NJ: Princeton University Press.

Columbus, G. 2010. A comparative analysis of invariant tags in three varieties of English. *English World-Wide* 31(3): 288–310.

Comrie, B. 1998. South Asian languages and linguistic typology. In *The Yearbook of South Asian Languages and Linguistics*, R. Singh (ed.), 237–246. New Delhi: Sage.

Constitution of India. <http://india.gov.in/govt/constitutions_india_bak.php#eng> (17 May 2011).

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Crystal, D. 2008. Two thousand million? Update on the statistics of English. *English Today* 24(1): 3–6.

Dasgupta, J. 2003. Language Policy and National Development in India. In *Fighting Words: Language Policy and Ethnic relations in Asia*, M.E. Brown & S. Ganguly (eds), 21–50. Cambridge, MA: MIT Press.

Dasgupta, P. 1993. *The Otherness of English: India's Auntie Tongue Syndrome*. New Delhi: Sage.

Drubig, H.B. & W. Schaffar. 2001. Focus Constructions. In. *Language Typology and Language Universals*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1079–104. Berlin: Mouton de Gruyter.

D'Souza, J. 1988. Interactional strategies in South Asian languages: Their implications for teaching English internationally. *World Englishes* 7(2): 159–171.

D'Souza, J. 1991a. Echoes in a sociolinguistic area. *Language Sciences* 13(2): 289–299.
D'Souza, J. 1991b. Speech acts in Indian English fiction. *World Englishes* 10(3): 307–316.
D'Souza, J. 1992. The relationship between code-mixing and the new varieties of English: Issues and implications. *World Englishes* 11(2–3): 217–223.
D'Souza, J. 1997. Indian English: Some myths, some realities. *English World-Wide* 18(1): 91–105.
D'Souza, J. 1999. Afterword. *World Englishes* 18(2): 271–274.
D'Souza, J. 2001. Contextualizing range and depth in Indian English. *World Englishes* 20(2): 145–159.
DeGraff, M. 2005. Linguists' most dangerous myth: The fallacy of Creole exceptionalism. *Language in Society* 34(4): 533–591.
Detarami, C. & Lock, G. 2003. Multilingualism in decline: Language repertoire, use and shift in two Hong Kong Indian communities. *Journal of Multilingual and Multicultural Development* 24(4): 249–273.
Dimock, E.C. Jr., Kachru, B.B. & Krishnamurti, B. (eds). 1992. *Dimensions of Sociolinguistics in South Asia*, 47–69. New Delhi: Oxford & IBH.
Ebert, K. 2001. Südasien als Sprachbund. In *Language Typology and Language Universals. An International Handbook*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1529–1539. Berlin: Mouton de Gruyter.
Emeneau, M.B. 1956. India as a linguistic area. *Language* 32(1): 3–16.
Ferguson, C.A. 1996. South Asia as a sociolinguistic area. In *Sociolinguistic Perspectives: Papers on Language in Society*; 1959–1994, C.A. Ferguson (ed.), 84–96. New York: Oxford University Press.
Filppula, M. 2006. The making of Hiberno-English and other "Celtic Englishes". In *The Handbook of the History of English*, A. van Kemenade & B. Los (eds), 507–536. Oxford: Blackwell.
Fuchs, R. 2012. Focus marking and semantic transfer in Indian English: The case of *also*. *English World-Wide* 33(1): 27–53.
Gandhi, M.K. 1921. Young India, 9.2.1921. In *The Collected Works of Mahatma Gandhi* Vol. 22. <http://www.gandhiserve.org/cwmg/cwmg.html>
Gardner-Chloros, P. 2009. *Code-Switching*. Cambridge: CUP.
Givón, T. 1979a. From discourse to syntax: Grammar as a processing strategy. In *Syntax and Semantics*, Vol. 12: *Discourse and Syntax*, T. Givón (ed.), 81–112. New York NY: Academic Press.
Givón, T. 1979b. *On Understanding Grammar*. New York NY: Academic Press.
Givón, T. 1993. *English Grammar: A Function-Based Introduction,* I & II. Amsterdam: John Benjamins.
Greenbaum, S. 1990. Standard English and the International Corpus of English. *World Englishes* 9: 79–83.
Greenbaum, S. (ed.). 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.
Gregory, M.L. & Michaelis, L.A. 2001. Topicalization and left-dislocation: A functional opposition revisited. *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies* 33(11): 1665–1706.
Gumperz, J.J. & Cook-Gumperz, J. 2005. Making space for bilingual communicative practice. *Intercultural Pragmatics* 2(1): 1–23.
Gumperz, J. & Wilson, R. 1971. Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In *Pidginization and Creolization of Languages*, D. Hymes (ed.), 151–167. London: CUP.

Gundel, J.K. 1988. Universals of topic-comment structure. In *Studies in Syntactic Typology* [Typological Studies in Language 17], M. Hammond, E. Moravcsik & J. Wirth (eds), 209–239. Amsterdam: John Benjamins.

Haugen, E. 1966. Dialect, language, nation. *American Anthropologist* 68: 922–35.

Heine, B. 2008. Contact-induced word order change without word order change. In *Language Contact and Contact Languages* [Hamburg Studies on Multilingualism 7], P. Siemund & N. Kintana (eds), 33–60. Amsterdam: John Benjamins.

Heine, B. & Kuteva, T. 2005. *Language Contact and Grammatical Change*. Cambridge: CUP.

Holmes, J. 1995. *Women, Men and Politeness*. London: Longman.

Hosali, P. 2008. Butler English: Morphology and syntax. In *Varieties of English*, 4: *Africa, South and Southeast Asia*, R. Mesthrie (ed.), 563–577. Berlin: Mouton de Gruyter.

Huddleston, R.D. & Pullum, G.K. (eds). 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Hudson-Ettle, D.M. & Schmied, J. 1999. *Manual to Accompany the East African Component of the International Corpus of English*. Chemnitz: Department of English, Chemnitz University of Technology.

Hundt, M. 2006. The committee has/have decided ... On concord patterns with collective nouns in inner- and outer-circle varieties of English. *Journal of English Linguistics* 34(3): 206–232.

Hundt, M. 2007. Global feature – local norms? A case study on the progressive passive". In *World Englishes – Problems, Properties and Prospects* [Varieties of English around the World G40], T. Hoffmann & L. Siebers (eds), 287–308. Amsterdam: John Benjamins.

Jain, D. 2003. Sociolinguistics of the Indo-Aryan languages. In *The Indo-Aryan Languages*, G. Cardona & D. Jain (eds), 46–66. London: Routledge.

Jayaram, B.D. & Rajyashree, K.S. (eds). 1998. *Goals and Strategies of Development of Indian Languages*. Mysore: Central Institute of Indian Languages.

Jespersen, O. 1969 [1937]. *Analytic Syntax.* New York NY: Holt, Rinehart and Winston.

Johanson, L. 2002. Contact-induced change in a code-copying framework. In *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, M.C. Jones & E. Esch (eds), 285–313. Berlin: Mouton de Gruyter.

Johanson, L. 2008. Remodeling grammar: Copying, conventionalization, grammaticalization. In *Language Contact and Contact Languages* [Hamburg Studies on Multilingualism 7], P. Siemund & N. Kintana (eds), 61–80. Amsterdam: John Benjamins.

Junghare, I.Y. 1988. Topic-Prominence and Zero NP-Anaphora in Marathi and Hindi. In *Languages & Cultures: Studies in Honour of Edgar C. Polomé*, M.A. Jazayery & W. Winter (eds), 309–27. Berlin: Mouton de Gruyter.

Kachru, B.B. 1986. *The Alchemy of English: The Spread, Functions, and Models of Non-Native Englishes*. Chicago IL: University of Illinois Press.

Kachru, B.B. 1988. The sacred cows of English. *English Today* 16: 3–8.

Kachru, B.B. 1991. Liberation linguistics and the Quirk concern. *English Today* 25: 3–13.

Kachru, B.B. 1994. English in South Asia. In *The Cambridge History of the English Language,* Vol. 5: *English in Britain and Overseas: Origins and Developments*, R. Burchfield (ed.), 497–553. Cambridge: CUP.

Kachru, B.B. 1998. English as an Asian language. *Links & Letters* 5: 89–108.

Kachru, B.B. 2005. *Asian Englishes. Beyond the Canon*. Hong Kong: Hong Kong University Press.

Kachru, B.B., Greenbaum, S. & Svartvik, J. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H.G. Widdowson (eds), 11–34. Cambridge: CUP.

Kachru, Y. 2006. *Hindi* [London Oriental and African Language Library 12]. Amsterdam: John Benjamins.

Kaviraj, S. 1992. Writing, speaking, being: language and the historical formation of identities in India. In *Nationalstaat und Sprachkonflikte in Süd- und Südostasien. Beiträge zur Südasienforschung, Südasien-Institut Universität Heidelberg*, D. Hellmann-Rajanayagam & D. Rothermund (eds), 25–68. Stuttgart: Franz Steiner Verlag.

Khubchandani, L.M. 1991. India as a sociolinguistic area. *Language Sciences* 13(2): 265–288.

King, R.D. 2008. Language politics and conflicts in South Asia. In *Language in South Asia*, B.B. Kachru, Y. Kachru & S.N. Sridhar (eds), 311–324. Cambridge: CUP.

Kiss, K.É. 2001. Discourse configurationality. In *Language Typology and Language Universals*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1445–1455. Berlin: Mouton de Gruyter.

Klippel, F. & Mukherjee, J. 2007. Standards and norms in language description and language teaching: An introduction. In *Anglistentag 2006: Proceedings*, S. Volk-Birke & J. Lippert (eds), 303–06. Trier: Wissenschaftlicher Verlag.

Koch, P. 1988. Norm und Sprache. In *Energeia und Ergon: Sprachliche Variation – Sprachgeschichte – Sprachtypologie*, Band II: *Das sprachtheoretische Denken Eugenio Coserius in der Diskussion*, H. Thun (ed.), 327–354. Tübingen: Narr.

Koch, P. & Oesterreicher, W. 1996. Sprachwandel und Expressive Mündlichkeit. *Zeitschrift für Literaturwissenschaft und Linguistik* 26(102): 64–96.

Koch, P. & Oesterreicher, W. 2012. Language of immediacy – language of distance: Orality and literacy from the perspective of language theory and linguistic history. In *Communicative Spaces: Variation, Contact, Change. Papers in Honour of Ursula Schaefer*, C. Lange, B. Weber & G. Wolf (eds), 441–473. Frankfurt: Peter Lang.

König, E. 1991. *The Meaning of Focus Particles. A Comparative Perspective*. London: Routledge.

König, E. 1993. Focus particles. In *Syntax. Ein Internationales Handbuch Zeitgenössischer Forschung/An International Handbook of Contemporary Research,* Vol. 1, J. Jacobs, A. von Stechow, W. Sternefeld & T. Vennemann (eds), 978–87. Berlin: Mouton de Gruyter.

König, E. & Siemund, P. 1999. Intensifiers and reflexives: A typological perspective. In *Reflexives. Forms and Functions* [Typological Studies in Language 40–41], Z. Frajzyngier & T.S. Curl (eds), 41–74. Amsterdam: John Benjamins.

Kortmann, B. & Szmrecsanyi, B. 2004. Global synopsis: Morphological and syntactic variation in English. In *A Handbook of Varieties of English,* Vol. 2: *Morphology and Syntax*, B. Kortmann, K. Burridge, R. Mesthrie, E.W. Schneider & C. Upton (eds), 1143–1202. Berlin: Mouton de Gruyter.

Kortmann, B., Burridge, K., Mesthrie, R., Schneider, E.W. & Upton, C. (eds). 2004. *A Handbook of Varieties of English*, Vol. 2: *Morphology and Syntax*. Berlin: Mouton de Gruyter.

Kortmann, B. & Lunkenheimer, K. (eds). 2011. *The electronic World Atlas of Varieties of English* [eWAVE]. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://www.ewave-atlas.org> (25 March 2012).

Koul, O.N. 1990. The use of the particles in Hindi. *International Journal of Dravidian Linguistics* 19(1): 22–36.

Krishnamurti, B. 2003. *The Dravidian Languages*. Cambridge: CUP.

Krishnaswamy, N. & Burde, A.S. 1998. *The Politics of Indians' English. Linguistic Colonialism and the Expanding English Empire*. New Delhi: OUP.

Krug, M. 1998. British English is developing a new discourse marker, *innit*? A study in lexicalisation based on social, regional and stylistic variation. *Arbeiten aus Anglistik und Amerikanistik* 23(2): 145–197.

Labov, W. 1972a. *Sociolinguistic Patterns*. Philadelphia PA: University of Pennsylvania Press.
Labov, W. 1972b. Some principles of linguistic methodology. *Language in Society* 1: 97–120.
Labov, W. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(2): 205–254.
Labov, W. 2001. *Principles of Linguistic Change: Social Factors*. Oxford: Blackwell.
Lambrecht, K. 1994. *Information Structure and Sentence Form. Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: CUP.
Lambrecht, K. 2001a. A framework for the analysis of cleft constructions. *Linguistics* 39(3): 463–516.
Lambrecht, K. 2001b. Dislocation. In *Language Typology and Language Universals,* M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1050–1078. Berlin: Mouton de Gruyter.
Lange, C. 2007a. Focus marking in Indian English. *English World-Wide* 28(1): 89–118.
Lange, C. 2007b. Let's face the music: The multilingual challenge. In *Annual Review of South Asian Languages and Linguistics*, R. Singh (ed.), 85–93. Berlin: Mouton de Gruyter.
Lange, C. 2009. 'Where's the party *Yaar*': Discourse markers in Indian English. In *World Englishes: Problems – Properties – Prospects* [Varieties of English around the World G40], T. Hoffmann & L. Siebers (eds), 207–225. Amsterdam: John Benjamins.
Lange, C. 2010. 'Hindi never, English ever' – Language nationalism and linguistic conflicts in modern India. In *Linguistics. Ideology and the Discourse of Linguistic Nationalism*, U. Schaefer, C. Lange & G. Wolf (eds), 133–154. Frankfurt: Peter Lang.
Lange, C. & Schaefer, U. 2008. "'Tis he, 'Tis she, 'Tis me, 'Tis – I don't know who..." Cleft and identificational constructions in 16th to 18th century English plays. In *Selected Papers from the Fourteenth International Conference on English Historical Linguistics (Icehl 14), Bergamo, 21–25 August 2006,* Vol. I: *Syntax and morphology* [Current Issues in Linguistic Theory 295], M. Gotti, M. Dossena & R. Dury (eds), 203–21. Amsterdam: John Benjamins.
Lass, R. 1980. *On Explaining Language Change*. Cambridge: CUP.
Le Page, R.B. 1988. Some premises concerning the standardization of languages, with special reference to Caribbean Creole English. *International Journal of the Sociology of Language* 71: 25–36.
Lehmann, T. 1993. *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
Li, C.N. & Thompson, S.A. 1976. Subject and topic: A new typology of language. In *Subject and Topic*, C.N. Li (ed.), 457–489. New York NY: Academic Press.
Lim, L. & Gisborne, N. 2009. The typology of Asian Englishes: Setting the agenda. *English World-Wide* 30(2): 123–132.
Lim, L. (ed.). 2004. *Singapore English. A Grammatical Description* [Varieties of English Around the World G33] Amsterdam: John Benjamins.
Lim, L. 2007. Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes* 26(4): 446–473.
Mahboob, A. 2008. Pakistani English: Morphology and syntax. In *Varieties of English 4: Africa, South and Southeast Asia*, R. Mesthrie (ed.), 578–592. Berlin: Mouton de Gruyter.
Mair, C. 2003. Kreolismen und verbales Identitätsmanagement im geschriebenen Jamaikanischen English. In *Zwischen Ausgrenzung und Hybridisierung*, E. Vogel, A. Napp & W. Lutterer (eds), 79–96. Würzburg: Ergon.
Martínez-Insua, A.E. 2004. *Existential* There-*Constructions in Contemporary British English: A Corpus-Driven Analysis of Their Use in Speech and Writing*. Munich: Lincom.

Masica, C.P. 1976. *Defining a Linguistic Area: South Asia*. Chicago IL: University of Chicago Press.

Masica, C.P. 1993. *The Indo-Aryan Languages*. Cambridge: CUP.

Masica, C.P. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In *The Yearbook of South Asian Languages and Linguistics. Tokyo Symposium on South Asian Languages: Contact, Convergence and Typology*, R. Singh (ed.), 205–267. New Delhi: Sage.

Matras, Y. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2): 281–331.

Matras, Y. 2000a. Mixed languages: A functional-communicative approach. *Bilingualism: Language and Cognition* 3(2): 79–99.

Matras, Y. 2000b. Fusion and the cognitive basis for bilingual discourse markers. *International Journal of Bilingualism* 4: 505–528.

Matras, Y. 2000c. How predictable is contact-induced change in grammar? In *Time Depth in Historical Linguistics*, C. Renfrew, A. McMahon & L. Trask (eds), 563–583. Oxford: The McDonald Institute for Archaeological Research.

Matras, Y. 2007a. Contact, connectivity and language evolution. In *Connectivity in Grammar and Discourse* [Hamburg Studies on Multilingualism 5], J. Rehbein, C. Hohenstein & L. Pietsch (eds), 51–74. Amsterdam: John Benjamins.

Matras, Y. 2007b. The borrowability of structural categories. In *Grammatical Borrowing in Cross-Linguistic Perspective*, Y. Matras & J. Sakel (eds), 31–73. Berlin: Mouton de Gruyter.

Matras, Y. 2009. *Language Contact*. Cambridge: CUP.

Matras, Y. & Sakel, J. 2007. Investigating the mechanisms of pattern replication in language convergence. *Studies in Language* 31(4): 829–865.

McArthur, T. 2003. *Oxford Guide to World English*. Oxford: OUP.

Mehrotra, R.R. 2003. A British response to some Indian English usages. *English Today* 75(19, 3): 19–25.

Mesthrie, R. 1992. *English in Language Shift. The History, Structure and Sociolinguistics of South African Indian English.* Cambridge: CUP.

Mesthrie, R. 1997. A sociolinguistic study of topicalisation phenomena in South African Black English. In *Englishes around the World*, I: *General Studies, British Isles, North America*; II: *Caribbean, Africa, Asia, Australia* [Varieties of English around the World G18–19] E.W. Schneider (ed.), 119–140. Amsterdam: John Benjamins.

Mesthrie, R. 2003. The World Englishes paradigm and contact linguistics: Refurbishing the foundations. *World Englishes* 22(4): 449–461.

Mesthrie, R. 2006a. World Englishes and the multilingual history of English. *World Englishes* 25(3–4): 381–390.

Mesthrie, R. 2006b. Contact linguistics and World Englishes. In *The Handbook of World Englishes*, B.B. Kachru, Y. Kachru & C.L. Nelson (eds), 273–288. Oxford: Blackwell.

Mesthrie, R. 2008a. English circling the globe: A commentary on a seminal ET article by Braj Kachru. *English Today* 93 (24, 1): 28–32.

Mesthrie, R. 2008b. Synopsis: Morphological and syntactic variation in Africa and South and Southeast Asia. In *Varieties of English*, 4: *Africa, South and Southeast Asia*, R. Mesthrie (ed.), 624–635. Berlin: Mouton de Gruyter.

Mesthrie, R. & Bhatt, R.M. 2008. *World Englishes. The Study of New Linguistic Varieties*. Cambridge: CUP.

Milroy, J. 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* 5(4): 530–555.

Mitchell, L. 2005. Parallel languages, parallel cultures: Language as a new foundation for the reorganisation of knowledge and practice in Southern India. *Indian Economic and Social History Review* 42(4): 445–467.

Moag, R.F. & Poletto, R. 1991. Discourse level evidence for South Asia as a linguistic area. *Language Sciences* 13(2): 229–254.

Montaut, A. 2005. Colonial language classification, post-colonial language movements, and the grassroots multilingualism ethos in India. In *Living Together Separately: Cultural India in History and Politics*, M. Hasan & A. Roy (eds), 75–116. New Delhi: OUP.

Mufwene, S. 1998. Native speaker, proficient speaker and norms. In *The Native Speaker: Multilingual Perspectives*, R. Singh (ed.), 111–123. New Delhi: Sage.

Mufwene, S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.

Mufwene, S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.

Mukherjee, J. 2007. Steady states in the evolution of New Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35(2): 157–187.

Mukherjee, J. 2010. The development of the English language in India. In *The Routledge Handbook of World Englishes*, A. Kirkpatrick (ed.), 167–180. London: Routledge.

Mukherjee, J. & Gries, S.T. 2009. Collostructional nativisation in New Englishes. *English World-Wide* 30(1): 27–51.

Mukherjee, J. & Hoffmann, S. 2006. Describing verb-complementation profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27(2): 147–173.

Myers-Scotton, C. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: OUP.

Nelson, G., Wallis, S. & Aarts, B. 2002. *Exploring Natural Language: Working with the British Component of the International Corpus of English* [Varieties of English around the World G29]. Amsterdam: John Benjamins.

Nelson, G. 1996. The design of the corpus. In *Comparing English Worldwide. The International Corpus of English*, S. Greenbaum (ed.), 27–35. Oxford: Clarendon Press.

Nelson, G. 1997. Cleft constructions in spoken and written English. *Journal of English Linguistics* 25(4): 340–348.

Netz, H. & Kuzar, R. 2007. Three marked theme constructions in Spoken English. *Journal of Pragmatics* 39: 305–335.

Nevalainen, T. & Tieken-Boon van Ostade, I. 2006. Standardisation. In *A History of the English Language*, R. Hogg & D. Denison (eds), 271–310. Cambridge: CUP.

Nevalainen, T. 1991. *But, Only, Just: Focusing Adverbial Change in Modern English 1500–1900* [Mémoires De La Société Néophilologique De Helsinki LI]. Helsinki: Société Néophilologique.

Nihalani, P., Tongue, R.K. & Hosali, P. 1979. *Indian and British English: A Handbook of Usage and Pronunciation*. New Delhi: OUP.

Odlin, T. 1997. Bilingualism and substrate influence: A look at clefts and reflexives. In *Focus on Ireland* [Varieties of Engish around the World G21], J. Kallen (ed.), 35–50. Amsterdam: John Benjamins.

Oesterreicher, W. 2001. Historizität – Sprachvariation, Sprachverschiedenheit, Sprachwandel. In *Language Typology and Language Universals*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 1554–1595. Berlin: Mouton de Gruyter.

Olavarría de Ersson, E. & Shaw, P. 2003. Verb complementation patterns in Indian Standard English. *English World-Wide* 24(2): 137–161.

Oxford English Dictionary online. <http://www.oed.com/> (17 May 2011)

Paikeday, T.M. 1985. *The Native Speaker Is Dead! An Informal Discussion of a Linguistic Myth with Noam Chomsky and Other Linguists, Philosophers, Psychologists, and Lexicographers*. Toronto: Paikeday.

Pandharipande, R. 1992. Defining politeness in Indian English. *World Englishes* 11(2–3): 242–250.

Pandharipande, R. 1997. *Marathi*. London, New York: Routledge.

Parasher, S.V. 1983. Indian English: Certain grammatical, lexical and stylistic features. *English World-Wide* 4(1): 27–42.

Parasher, S.V. 1991. *Indian English: Functions and Form*. New Delhi: Bahri Publications.

Pattanayak, D.P. 1998. Mother tongue: An Indian context. In *The Native Speaker: Multilingual Perspectives*, R. Singh (ed.), 127–147. New Delhi: Sage.

Platt, J., Weber, H. & Lian, H.M. 1984. *The New Englishes*. London: Routledge and Kegan Paul.

Pollock, S. 2006. *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India*. Berkeley CA: University of California Press.

Poplack, S. & Sankoff, D. 1984. Borrowing: The synchrony of integration. *Linguistics* 22: 99–135.

Poplack, S. & Meechan, M. 1998. How languages fit together in codemixing. *International Journal of Bilingualism* 2(2): 127–138.

Poplack, S. & Levey, S. 2010. Contact-induced grammatical change: A cautionary tale. In *Language and Space: An International Handbook of Linguistic Variation,* Vol. I: *Theories and Methods*, P. Auer & J.E. Schmidt (eds), 391–419. Berlin: Mouton de Gruyter.

Poplack, S. 1980. 'Sometimes I'll start a sentence in Spanish Y termino en Español': Toward a typology of code-switching. *Linguistics* 18: 581–618.

Poplack, S. 2004. Code-switching/Sprachwechsel. In *Sociolinguistics/Soziolinguistik: An International Handbook of the Science of Language and Society/Ein Internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft* 2nd rev. edn, U. Ammon, N. Dittmar, K.J. Matheier & P. Trudgill (eds), 589–596. Berlin: Mouton de Gruyter.

Primus, B. 2001. Word order and information structure: A performance-based account of topic positions and focus positions. In *Language Typology and Language Universals*, M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds), 880–896. Berlin: Mouton de Gruyter.

Prince, E.F. 1997. On the functions of left-dislocation in English discourse. In *Directions in Functional Linguistics* [Studies in Language Companion Series 36] A. Kamio (ed.), 117–143. Amsterdam: John Benjamins.

Prince, E.F. 1984. Topicalization and left-dislocation: A functional analysis. In *Discourses in Reading and Linguistics*, S.J. Teller & V. White (eds), 213–25. New York: New York Acad. of Sciences.

Quirk, R., Kennedy, G. & Crystal, D. 1985. The English language in a global context. In *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk & H.G. Widdowson (eds), 1–10. Cambridge: CUP.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Quirk, R. 1990. Language varieties and standard language. *English Today* 21: 3–10.

Rai, A. 1984. *A House Divided: The Origin and Development of Hindi/Hindavi*. Delhi: OUP.

Rogers, C. 2003. Register Variation in Indian English. PhD dissertation, Northern Arizona University.

Romaine, S. 2006. Global English: From island tongue to world language. In *The Handbook of the History of English*, A. van Kemenade & B. Los (eds), 589–608. Oxford: Blackwell.

Sahgal, A. & Agnihotri, R.K. 1985. Syntax – The common bond. Acceptability of syntactic deviances in Indian English. *English World-Wide* 6(1): 117–129.

Sailaja, P. 2009. *Indian English*. Edinburgh: EUP.

Sand, A. 2004. Shared morpho-syntactic features in contact varieties of English: Article use. *World Englishes* 23(2): 281–298.

Schaefer, U. 2006. The beginnings of standardization: The communicative space in fourteenth-century England. In *The Beginnings of Standardization: Language and Culture in Fourteenth-Century England*, U. Schaefer (ed.), 3–24. Frankfurt: Peter Lang.

Schaefer, U. 2010. Linguistics, ideology and the discourse of linguistic nationalism: Some preliminary remarks. In *Linguistics, Ideology and the Discourse of Linguistic Nationalism*, C. Lange, U. Schaefer & G. Wolf (eds), 1–36. Frankfurt: Peter Lang.

Schiffman, H. 1983. *A Reference Grammar of spoken Kannada*. Seattle WA: University of Washington Press.

Schilk, M. 2006. Collocations in Indian English: A corpus-based sample analysis. *Anglia* 124(2): 276–316.

Schilk, M. 2011. *Structural Nativization in Indian English Lexicogrammar* [Studies in Corpus Linguistics 46]. Amsterdam: John Benjamins.

Schmied, J. 1994. Syntactic style variation in Indian English. In *Anglistentag 1993 Eichstätt: Proceedings*, G. Blaicher & B. Glaeser (eds), 217–231. Tübingen: Max Niemeyer.

Schmied, J. 1996. Second-language corpora. In *Comparing English Worldwide: The International Corpus of English*, S. Greenbaum (ed.), 182–196. Oxford: Clarendon Press.

Schneider, E.W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281.

Schneider, E.W. 2004. How to trace structural nativization: Particle verbs in World Englishes. *World Englishes* 23(2): 227–249.

Schneider, E.W. 2007a. Editor's report 2003–2007. *English World-Wide* 28(3): 349–355.

Schneider, E.W. 2007b. *Postcolonial English. Varieties around the World*. Cambridge: CUP.

Sedlatschek, A. 2007. Variation and Change in Contemporary Indian English: A Corpus-Based Description. PhD dissertation, University of Freiburg.

Sedlatschek, A. 2009. *Contemporary Indian English: Variation and Change* [Varieties of English Around the World G38]. Amsterdam: John Benjamins.

Sharma, D. 2003. Structural and Social Constraints on Non-Native Varieties of English. PhD dissertation, Stanford University.

Sharma, D. 2001. The pluperfect in native and non-native English: A comparative corpus study. *Language Variation and Change* 13: 343–373.

Sharma, D. 2003. Discourse clitics and constructive morphology in Hindi. In *Nominals: Inside and Out*, M. Butt & T. Holloway King (eds), 59–84. Stanford CA: CSLI.

Sharma, D. 2005a. Dialect stabilization and speaker awareness in non-native varieties of English. *Journal of Sociolinguistics* 9(2): 194–224.

Sharma, D. 2005b. Language transfer and discourse universals in Indian English article use. *Studies in Second Language Acquisition* 27(4): 535–566.

Sharma, D. 2009. Typological diversity in new Englishes. *English World-Wide* 30(2): 170–195.

Shastri, S.V. 1988. The Kolhapur Corpus of Indian English and work done on its basis so far. *ICAME Journal* 12: 15–26.

Shastri, S.V. 2002. Overview of the Indian component of the International Corpus of English (ICE-India).

Sheorey, R. & Bhaskaran Nayar, P. 2002. Learning and teaching English in India: Looking from the outside. *Indian Journal of Applied Linguistics* 28(2): 13–24.

Singh, R.P. (ed.). 2004. *Language Atlas of* India: 1991, *Census of India*. Delhi: Controller of Publications.

Singh, R. 2003. The languages of India: A bird's-eye-view. In *The Yearbook of South Asian Languages and Linguistics*, R. Singh (ed.), 173–181. Berlin: Mouton de Gruyter.

Singh, R. 2006. Native speaker. In *Encyclopedia of Language and Linguistics, Second Edition*, Vol. 8, K. Brown (ed.), 489–492. Amsterdam: Elsevier.

Singh, R. 2007. The nature, structure, and status of Indian English. In *Annual Review of South Asian Languages and Linguistics 2007*, R. Singh (ed.), 33–46. Berlin: Mouton de Gruyter.

Sridhar, K.K. 1985. Sociolinguistic theory and non-native varieties of English. *Lingua* 68: 39–58.

Sridhar, S.N. 1992. The ecology of bilingual competence: Language interaction in indigenized varieties of English. *World Englishes* 11(2–3): 141–150.

Sridhar, S.N. 1996. Toward a syntax of South Asian English: Defining the lectal range. In *South Asian English: Structure, Use, and Users*, R.B. Baumgardner (ed.), 55–69. Urbana IL: Illinois University Press.

Subbarao, K.V., Agnihotri, R.K. & Mukherjee, A. 1991. Syntactic strategies and politeness phenomena. *International Journal of the Sociology of Language* 92: 35–53.

Subbarao, K.V. 2008. Typological characteristics of South Asian languages. In *Language in South Asia*, B.B. Kachru, Y. Kachru & S.N. Sridhar (eds), 49–78. Cambridge: CUP.

Tagliamonte, S.A. & D'Arcy, A. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85: 58–108.

Talbot, M. 2003. Gender stereotypes: Reproduction and challenge. In *The Handbook of Language and Gender*, J. Holmes & M. Meyerhoff (eds), 468–486. Oxford: Blackwell.

Thomason, S.G. 2001. *Language Contact: An Introduction*. Washington DC: Georgetown University Press.

Thomason, S.G. 2003. Contact as a source of language change. In *The Handbook of Historical Linguistics*, B.D. Joseph & R. Janda (eds), 687–712. Malden MA: Blackwell.

Thomason, S.G. 2008. Social and Linguistic Factors as Predictors of Contact-Induced Change. *Journal of Language Contact* THEMA 2: 42–56. </http: www.jlc-journal.org/>

Thomason, S.G. & Kaufman, T. 1991[1988]. *Language Contact, Creolization, and Genetic Linguistics*, 2nd edn. Berkeley CA: University of California Press.

Tottie, G. & Hoffmann, S. 2006. Tag questions in British and American English. *Journal of English Linguistics* 34(4): 283–311.

Trautmann, T.R. 2006. *Languages and Nations. The Dravidian Proof in Colonial Madras*. Berkeley CA: University of California Press.

Trudgill, P. & Hannah, J. 2002. *International English: A Guide to Varieties of Standard English*. London: Arnold.

Vaish, V. 2005. A peripherist view of English as a language of decolonization in post-colonial India. *Language Policy* 4: 187–206.

Valentine, T. 1988. Developing discourse types in non-native English: Strategies of gender in Hindi and Indian English. *World Englishes* 7(2): 143–158.

Valentine, T. 1991. Getting the message across: Discourse markers in Indian English. *World Englishes* 10(3): 325–334.

Valentine, T. 1995. Agreeing and disagreeing in Indian English discourse: Implications for language teaching. In *Language and Culture in Multilingual Societies: Viewpoints and Visions*, M.L. Tickoo (ed.), 227–250. Singapore: SEAMEO Regional Language Centre.

Van Coetsem, F. 2000. *A General and Unified Theory of the Transmission Process in Language Contact*. Heidelberg: Winter.

Van Marle, J. 1997. Dialect versus standard language: Nature versus culture. In *Taming the Vernacular: From Dialect to Written Standard Language*, J. Cheshire (ed.), 13–34. London: Longman.

Verma, M.K. 2002. English in India: Whose English, for whom, and what about Indian languages? *Indian Journal of Applied Linguistics* 28(2): 101–120.

wa Thiong'o, N. 1994. *Decolonising the Mind: The Politics of Language in African Literature*. London: Currey.

Ward, G.L. & Prince, E.F. 1991. On the topicalization of indefinite NPs. *Journal of Pragmatics* 16(2): 167–177.

Ward, G. & B. Birner. 2004. Information structure and non-canonical syntax. In T*he Handbook of Pragmatics*, L.R. Horn & G. Ward (eds.), 153–74. Malden, Oxford: Blackwell.

Washbrook, D. 1991. 'To each a language of his own': Language, culture and society in colonial India. In *Language, History and Class*, P. Corfield (ed.), 179–203. Oxford: Basil Blackwell.

Weinreich, U., Labov, W. & Herzog, M.I. 1968. Empirical foundations for a theory of language change. In *Directions for Historical Linguistics: A Symposium*, W.P. Lehmann & Y. Malkiel (eds), 97–195. Austin TX: University of Texas Press.

Weinreich, U. 1953. *Languages in Contact*. The Hague: Mouton.

Williams, J. 1987. Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8(2): 161–199.

Winford, D. 2005. Contact-induced changes: Classification and processes. *Diachronica* 22(2): 373–427.

Wright, R. 2004. Latin and English as world languages. *English Today* 80(20,4): 3–13.

Author index

Language index

Subject index