

IVÁN TAMAREDO

University of Santiago de Compostela – ivan.tamaredo@usc.es

The conventionalization of performance preferences: Pronoun omission in Indian English and Singapore English¹

1. Introduction

The focus of the present chapter is pronoun omission in Indian English and Singapore English, understood as the existence of a gap in a structure that could be filled by an overt personal pronoun, as in example (1):

- (1) B: *Do you know who she_i is*
A: *No Ø_i doesn't sound familiar* (ICE-SIN:S1A-092#243-244:1)

According to the *Electronic World Atlas of Varieties of English* (eWAVE) (Kortmann and Lunkenheimer 2013), a database comprising 235 grammatical features mapped onto 76 varieties of English, pronoun omission is attested in 51% and 37% of the varieties included in subject and direct object positions respectively. This fact could come as a surprise since English is usually considered a language that forbids the omission of arguments and, therefore, varieties of English could be expected to behave similarly in this respect.² However, as will be argued below,

1 This work was supported by the European Regional Development Fund, the Spanish Ministry of Economy and Competitiveness under grants FFI2014-51873-REDT, FFI2014-52188-P and BES-2015-071233, and the Directorate General for Scientific and Technological Promotion of the Regional Government of Galicia under grant GPC2014/004. I would like to thank J. Carlos Acuña-Fariña and Teresa Fanego for valuable comments on an earlier version of this article.

2 But see Huddleston and Pullum *et al.* (2002: 1540–1541) and Ruppenhofer and Michaelis (2010) for a description of certain contexts in which pronouns can be omitted in English.

pronoun omission renders linguistic structures more efficient and easier to process, a fact that makes it communicatively advantageous, especially for the speakers of certain types of varieties. A different issue is whether these preferences of the speakers are actually conventionalized as part of the grammars of varieties or if they are only a matter of performance or use. Ascertaining whether omitted pronouns are conventionalized anaphoric elements that refer to highly accessible antecedents in Indian English and/or Singapore English will be the main goal of the present study.

This chapter is structured as follows. Section 2 provides an overview of the concepts of complexity and communicative efficiency and describes how to measure them. Section 3 discusses the syntactic and pragmatic restrictions on the occurrence of omitted pronouns as identified in previous research. Section 4 examines the attestation of pronoun omission in different types of varieties of English, and this distribution is explained in terms of the amount of contact in the history of the speech communities in which the varieties are spoken. Section 5 deals with the methodology of a corpus study on the distribution of omitted and overt pronouns in Indian English and Singapore English. Sections 6 and 7 focus on the results of this study and their discussion, and, finally, section 8 presents some concluding remarks.

2. Complexity and efficiency

The concept of complexity has figured prominently in linguistics since the early days of the discipline. For the most part of the twentieth century, linguistic theorizing was guided by the assumption that languages are all equally complex because, even though they may differ in the complexity of specific grammatical domains (for example, morphology or syntax), there are always trade-offs between domains that compensate for possible differences and make all languages converge at some unspecified level of overall complexity (for an explicit formulation of this idea, see Hockett 1958: 180–181). Around the turn of the twenty-first century, a number of studies appeared that questioned this assumption

and devised ways of measuring grammatical complexity in an empirical manner (see, for instance, the collection of articles in Miestamo *et al.* 2008). The most widely held position nowadays is that languages may differ in overall complexity, and that trade-offs in complexity between different grammatical domains may occur but are not obligatory.

Several factors causing complexity differences between languages have been identified, but probably the one that has received most attention is language contact.³ A high number of second language users in the history of a speech community may lead to grammatical simplification “due to the relative inability of adult humans to learn new languages perfectly” (Trudgill 2009a: 99), which means that languages spoken in communities with a history of contact are simpler than the rest. This issue of simplification due to contact will be discussed further in section 4.

Before dealing with whether languages differ in complexity or they are all equal in this respect, it is important to state clearly how the notion of linguistic complexity is going to be operationalized and measured. First, we must distinguish between absolute and relative metrics of complexity (Miestamo 2008). On the one hand, in absolute metrics complexity is understood as an objective property of grammars, which are in turn conceptualized as abstract systems that are independent from considerations of language use. Relative metrics, on the other hand, are based on usage preferences: a grammar is complex if it has features that are difficult to learn or to process for language users. Another important distinction is that between systemic and structural metrics (Dahl 2004). Systemic metrics measure the complexity of the rules of a grammar that generate the linguistic structures used by speakers, while structural metrics measure the complexity of the actual structures that are the outputs of those rules. These distinctions are important because different types of complexity metrics provide different results.

In the present paper, Hawkins’ (2004) metric of communicative efficiency will be used to measure linguistic complexity. This is a relative and structural metric because it measures the processing difficulty of individual constructions. Hawkins’ (2004) metric integrates complexity into a

3 See Sampson (2009) for a review of factors that have been found to influence grammatical complexity.

broader notion that also takes into account the communicative function of language, that is, the transmission of information from the speaker to the hearer. Communication is considered to be efficient if the message that the speaker has in mind is conveyed to the hearer using structures that are easy to process and allow a fast transmission of information (Hawkins 2009: 253). Three different principles, formulated as processing preferences, are put forward by Hawkins (2004) in order to measure the efficiency of structures: Minimize Domains, Maximize On-line Processing, and Minimize Forms. Only the last principle, Minimize Forms, will be discussed in depth here since this is the one that is relevant for the analysis of the linguistic feature of pronoun omission. It is defined in the following terms:

The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) [...]. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F (Hawkins 2004: 38).

Simply put, structures that convey a message using fewer phonemes, words or phrases are more efficient and easier for the speaker to process and articulate, because processing linguistic forms and their associated meanings requires an effort. However, these minimal structures must also be interpretable, that is, the hearer must be able to decode the message from the forms that he or she receives. For this task, the hearer can resort to other sources of information, such as the linguistic and extra-linguistic contexts, and thus fill the gaps left in the structure by the speaker. This process is called “processing enrichment” (Hawkins 2004: 41), and it will be an important notion when discussing the complexity and efficiency of pronoun omission in the next section.

Finally, an important part of Hawkins’ theory is the “Performance Grammar Correspondence Hypothesis” (Hawkins 2004: 3). He proposes that speakers usually prefer to produce communicatively efficient structures and that these preferences are later reflected in the fixed conventions of grammar. This hypothesis, therefore, provides an account of language variation and change in which grammar is shaped by usage preferences. One of the goals of the present chapter is to test the Performance Grammar Correspondence Hypothesis with respect to pronoun omission in two varieties of English, Indian English and Singapore English.

3. Pronoun omission

The term pronoun omission is used here in order to refer to a gap in a structure that could be filled by a personal pronoun. Only those cases of omission of pronouns in subject and direct object positions will be considered.⁴ Examples (2) and (3), taken from the Indian (ICE-IND) and Singaporean (ICE-SIN) components of the International Corpus of English (ICE), are instances of omitted pronouns in Indian English and Singapore English in subject (2) and object (3) positions:

- (2) *The honourable minister has made a statement saying within three years [the requirements of the crude in this country] \emptyset_i have increased to <,> upto sixty percent of the last <,> pre previous production <,>* (ICE-IND:S1B-055#12:1:C)⁵
- (3) B: *I considered it_i*
A: *I don't care whether you considered \emptyset_i* (ICE-SIN:S1B-069#125-126:1)

Pronoun omission is a grammatical feature that has been widely analyzed cross-linguistically, mostly from a generative perspective. Most theories of grammar included in the generative tradition recognize the existence of empty categories, that is, elements that do not have phonetic realization but fulfil a syntactic role. Among these elements, there are two that are relevant to pronoun omission as understood here: pros and variables.⁶ The empty category *pro* behaves like an overt pronoun in

4 Cases of pronoun omission in subject position in non-finite clauses are not taken into account because they belong to a different class of empty categories (see below). Cases in imperative clauses are also excluded.

5 The NP in brackets is the antecedent of the omitted pronoun and it is located in the preceding discourse.

6 Traditionally, four different types of empty categories are distinguished: *NP-traces*, *pro*, *PRO* (usually the subject in non-finite clauses) and *variables*. Out of these, *pro* and *PRO* are generated in situ, while *NP-traces* and *variables* are left behind by movement operations (e.g. raising or topicalization) and are coindexed with the moved element. The antecedent of a *variable* moves to a nonthematic position, such as sentence topic (see example (5)), while that of a *NP-trace* moves to a thematic one, such as subject or direct object. See Haegeman (1994: 433–40, 450–4) for an account of the different types of empty categories distinguished in Government and Binding theory.

that it refers to an entity in the linguistic or extra-linguistic context. Not all languages allow the occurrence of *pro*; in some of them an overt element must be present. Verbal morphology plays an important role in explaining this variation between languages: *pro* is usually allowed in languages with rich subject/object-verb agreement because the referent of the empty element can be recovered from the affixes attached to the verb. In Spanish, for instance, subjects can be omitted because they agree in person and number with the verb:

- (4) *Ayer Ø_i vi_i a Juan y Ø_j me dijo_j que Ø_k habló_k con María*
 Yesterday see+1sing.PERF Juan and me tell+3sing.PERF that talk+3sing.PERF
 to María
 [I] saw Juan yesterday and [he] told me that [he] talked to María

However, as pointed out by Huang (1984), there are languages with poor or no agreement, such as Chinese, which allow omission both in subject and object positions. In these languages, according to Huang's analysis, only cases of omission in subject position in subordinate clauses that refer back to an entity in the matrix clause are instances of *pro*, as in (5):

- (5) *Zhangsan_i xiwang Ø_i keyi kanjian Lisi*
 Zhangsan hope can see Lisi
 Zhangsan_i hopes that [he_i] can see Lisi (Huang 1984: 538)

while the rest of the cases are variables coindexed with empty sentence topics, as in the following example:

- (6) [_{Top} Ø_i], *Zhangsan shuo Lisi bu renshi Ø_i*
 Zhangsan say Lisi not know
 *[Him]_i, Zhangsan said that Lisi didn't know (Huang 1984: 542)

In (6), the object of the subordinate clause is first topicalized and placed at the beginning of the sentence; then, this topic is omitted.⁷

7 A complete review of all the existing analyses of pronoun omission is beyond the purpose of this chapter (for such an overview see, for instance, Cole 2010), but it should be noted that Huang's account has been challenged by Xu (1986) and Bao (2001), among others. Xu argues that omitted objects in Chinese cannot be *variables* coindexed with empty topics because they do not comply with island constraints, which are restrictions on movement

Most descriptions of pronoun omission recognize that, despite the formulation of purely syntactic restrictions on the occurrence of omitted pronouns, a complete account of this phenomenon requires taking into account issues of a pragmatic nature. Cole (2010: 280) points out that, even in languages with rich agreement, this is not always enough to license the occurrence of an omitted pronoun due to syncretism (that is, when the same affix is used for different person-number combinations), and in these cases the referent must be retrieved from the context. Therefore, the presence of an antecedent in the linguistic or extra-linguistic context that is accessible for the hearer must be a component of any account of pronoun omission. There are different factors that influence the accessibility of referents in discourse but, following Ariel's (1990, 1994, 2001) Accessibility theory, it will be considered here that a referent is highly accessible when it is salient and when there is a tight link between the overt/omitted pronoun and its antecedent (see section 5 for further detail).

As noted in the previous section, structures are communicatively more efficient when they have fewer forms, provided that the hearer is able to decode the message. Pronoun omission, therefore, increases the efficiency of structures because it minimizes their formal complexity, making them easier to process. However, this is the case only if the referents of omitted pronouns are accessible in discourse, thus allowing structures to be enriched in processing and reducing the effort that the hearer has to make in order to interpret successfully the meaning conveyed by the speaker.⁸

operations. Bao claims that omitted objects in subordinate clauses can also refer to an element in the matrix clause, thus behaving like *pros*, not like *variables*.

8 It should also be mentioned, as an anonymous referee has noted, that the opposite trend is also found in contact situations: in certain cases, redundancy and the overt occurrence of elements that tend to be omitted may be necessary in order to prevent possible comprehension errors and thus keep communication efficient (see, for instance, Williams 1987, 1988; Schneider 2013).

4. Pronoun omission and varieties of English

As mentioned in section 2, language contact is a factor that affects grammatical complexity. In the present section, it will be shown that pronoun omission is a feature commonly found in the speech of second language users of English and in high-contact varieties of English, and it will be argued that this is because pronoun omission results in more efficient and easier-to-process structures.

Williams (1988) examined the distribution of omitted third person pronouns in subject position in the speech of native speakers of English, second language speakers of English with different first languages, and speakers of Singapore English, which is a high-contact variety. She found that second language and Singapore English speakers dropped pronouns more frequently and in a wider variety of contexts than native speakers of English, who mostly restricted pronoun omission to the second conjunct of explicit or implicit coordinate clauses, as in (7):

(7) [...] *He_i just went there at lunch time and Ø_i found out* (Williams 1988: 352)

Gundel and Tarone (1983) found that the omission of pronouns in direct object position is a characteristic feature of the speech of second language speakers of English with Spanish as a first language, even though the grammar of Spanish does not contemplate the omission of direct objects. These studies show that pronoun omission is a common feature in the production of second language learners, which is not surprising if we take into account that it makes structures more efficient and easier to process and that simplification seems to be a result of non-native language acquisition and use (see, for instance, Klein and Perdue 1997).

Pronoun omission is also commonly attested in high-contact varieties of English, that is, those spoken in communities with a high number of adult language learners (in the present or in the past). As mentioned in the introduction, eWAVE includes information about the attestation of pronoun omission in varieties of English. As shown in Table 1, omission is more frequently attested in high-contact varieties than in low-contact ones. This classification is based on Trudgill (2009b) with a minor modification: the high-contact category includes

indigenized second language varieties (e.g. Hong Kong English), shift varieties (e.g. Irish English), pidgins (e.g. Nigerian Pidgin) and creoles (e.g. Hawai'i Creole); the low-contact category comprises traditional first language dialects (e.g. Newfoundland English) and dialect-contact varieties (e.g. New Zealand English).

Table 1: Attestation of pronoun omission in varieties of English (eWAVE).

VARIETY TYPE	PRONOUN OMISSION	
	<i>Attested</i>	<i>Not Attested</i>
<i>Low-contact</i>	38.9% (7)	61.1% (11)
<i>High-contact</i>	72.4% (42)	27.7% (16)

Omitted pronouns occur in less than 40% of low-contact varieties but in more than 70% of high-contact ones, a finding that agrees with the facts mentioned above: pronoun omission simplifies structures and second language acquisition tends to simplify languages.

In section 2, Hawkins' Performance Grammar Correspondence Hypothesis was introduced. According to it, grammars reflect the preferences of users, and speakers tend to produce efficient structures whenever possible. Therefore, high-contact varieties of English may have conventionalized pronoun omission in their grammars due to the fact that second language speakers are prone to omit pronouns when their referents can be successfully recovered by hearers. The following sections will deal with a corpus study of two high-contact varieties, Indian English and Singapore English, in which this hypothesis is tested.

5. Materials and methodology

The aim of the present study is to examine the distribution of omitted and overt pronouns in Indian English and Singapore English with respect to the accessibility of their referents in order to ascertain

whether these two types of anaphoric elements are used to refer to different classes of antecedents. Moreover, if they are indeed distributed differently, a further goal is to establish whether these preferences are conventionalized in the grammars of the varieties under consideration. ICE-IND and ICE-SIN were used as a source of data. Indian English and Singapore English were chosen because they present a high frequency of pronoun omission, according to eWAVE. Additionally, both are high-contact varieties but they belong to different subtypes: Indian English is an indigenized second language variety, which means that the majority of its speakers are second language users of English (Lange 2012: 54), and Singapore English is becoming a shift first language, that is, a variety that was mainly a second language but that is now being acquired and used as a first language by a portion of the population (Leimgruber 2013: 7).

The concordance software WordSmith Tools was used in order to retrieve the instances of overt pronouns, but the instances of omitted pronouns had to be identified manually. The corpus that was analyzed for the present study comprises forty texts taken from ICE-IND and ICE-SIN (approximately ninety thousand words, forty-five thousand per variety). Of the forty texts selected, twenty belonged to ICE-IND and the other twenty to ICE-SIN. The texts were chosen evenly with respect to register and mode so that, for each variety, five texts were spoken and informal, five spoken and formal, five written and informal, and five written and formal. ICE provides a classification of texts regarding their genre, and this classification was followed in the present study: spoken informal texts were selected from the S1A category (face-to-face and telephone conversations), spoken formal texts from the S1B category (mainly from parliamentary debates and legal cross-examinations), written informal texts from the W1B category (from the genre of social letters) and, finally, written formal texts from the W2A (academic writing), W2C (news reports), W2D (instructional writing) and W2E (press editorials) categories.

The accessibility of referents was measured on the basis of Ariel's (1990, 1994, 2001) Accessibility theory, which was operationalized in the following terms.⁹ The accessibility of an entity or event increases if

9 Non-referential pronouns, such as the first occurrence of *it* in *So naturally it's very difficult for us to <,> take uh <,> things inside and capture it and then*

it is salient in the discourse and if there is a tight link between the anaphoric element and its linguistic antecedent.¹⁰ The saliency of the referent increases:

- If it is a discourse topic: an entity was considered to have high topicality if it was mentioned at least twice before the occurrence of the omitted/overt pronoun. It has been argued that only those mentions located up to four propositions to the left of the anaphoric element should be counted (Toole 1996), and that mentions in the previous clause should be considered more important than the rest (Clark and Sengul 1979). In the present study, however, all the occurrences of an entity in the text were taken into account in order to measure its topicality, and distance is treated as an independent factor.
- If it is emphasized, for instance, by being preposed, postposed, by being part of a cleft or a pseudo-cleft sentence, etc.
- If it is in subject position.
- If it is animate rather than inanimate.
- If it is not ambiguous, that is, if there are no other potential antecedents between the anaphoric element and the correct antecedent, or in the three clauses preceding it.
- If it was last coded by a low accessibility marker, for instance, a full noun phrase rather than a pronoun.

The link between the anaphoric element and its antecedent is tighter:

- If the anaphor and the antecedent are in embedded clauses.
- If they are in coordinate clauses.

produce it (ICE-IND:S1A-010#61:1:A), were excluded from the present study because they do not have antecedents (they only fulfil a structural function) and, therefore, cannot be analyzed in terms of the accessibility of their referents.

10 Those anaphoric elements that referred to the participants in the communicative exchange (mainly first and second person pronouns) were considered to have exophoric reference, that is, they refer directly to an entity in the extra-linguistic context and do not have linguistic antecedents. These cases, therefore, could only be analyzed with respect to the saliency of their referents and not to the tightness of the link between the anaphor and its antecedent.

- If the sentences containing them are pragmatically cohesive, that is, if they are connected by conjunctive adverbs or if the antecedent is in an initiating speech act (e.g. a question) and the anaphoric element in a responding speech act (e.g. the answer to the question).
- If there is a short distance in clauses – not more than four – between the anaphor and its antecedent. Both finite and non-finite clauses were taken into account, and utterances without a verb that were located in an independent turn were also counted, with the exception of back channel words such as *yeah* or *uhm*. There is no agreement in the literature with respect to how long an entity remains active in memory. Givón (1983) proposed an arbitrary limit of twenty clauses to the left of the anaphoric element. The present paper deals with markers of high accessibility (omitted and overt pronouns) and, therefore, only shorter distances are taken into consideration. Williams (1988) found that there is an average distance of three to four clauses between omitted pronouns and their antecedents in the speech of second language speakers of English and Singapore English speakers. For this reason, distances longer than four clauses are not considered here to increase the accessibility of the antecedent.

Three levels of accessibility were distinguished: high, intermediate, and low. Highly accessible referents were those that were salient and, if applicable, showed a tight link between the anaphoric element and its antecedent, as in example (8). The intermediate level comprised those referents that were either salient or had a tight link, as in (9). Finally, referents with low accessibility were those that were not salient and did not show a tight link between the anaphor and the antecedent, as in example (10).

- (8) A: *I don't know uhm because the the last time uhm Kim_i came back I bump into Kim_i in at the Andrew's shop*
 A: *So so already told me Ø_i said that he's going to be in San Francisco and he ask I think he's managed to get Kim uh Kong Shin's number from Lionel I think*
 (ICE-SIN:S1A-062#3-4:1)
- (9) A: *[...] would you like to go today <,> as a well qualified person <,> and go <,> and work in Shimoga <,> or some place called Shirshi <,> would you like [to work like that]_i*
 B: *They will not serve*

- (10) A: *You won't like \emptyset_i <, > (ICE-IND:S1A-045#133–135:1)¹¹*
 C: *So you can bring it_i <, > it may be good if you can record it [...]*
 B: *Yeah <, > but only I wanted your permission*
 C: *No it'll be a welcome <, > uhm <, > because university is saying this thing is not working <, > ah <, > so if you <, > maybe come <, > a little bit earlier <, > yeah <, > then they can plug \emptyset_i in and <, > test it <, > yeah <, > because you <, > that mike condition system we don't know (ICE-IND:S1A-095#68–73:2)¹²*

6. Results

After the retrieval process, 205 and 286 instances of omitted pronouns were identified in Indian English and Singapore English respectively. The number of overt pronouns retrieved was 2682 in Indian English and 2884 in Singapore English. This means that, as shown in Table 2, the total frequency of pronoun omission is 7.1% in Indian English and 9% in Singapore English, which indicates that omitted pronouns are not particularly widespread in either variety. There is a higher frequency of omission in Singapore English than in Indian English, and this distribution is statistically significant ($\chi^2(1) = 7.49, p < .01$).¹³ The percentage of omission is lower in subject position in Indian English (6.3%) than in Singapore English (9.2%), but it is higher in direct object position (12.9% versus 7.3%), and this distribution is also statistically significant ($\chi^2(4) = 30.14, p < .001$).

11 There are other potential antecedents in the immediate preceding context (the cities of Shimoga and Shirsi), which decrease the saliency of the referent, but there is a short distance between the omitted pronoun and its antecedent, which increases the tightness of their link.

12 The antecedent *it* refers to a tape recorder. There are other potential antecedents (the pronoun *it* in *if you can record it*, which refers to a speech that the participants of the conversation want to record, or *this thing*), which decrease the saliency of the referent. Furthermore, the omitted pronoun and the antecedent are separated by more than four clauses.

13 The test chosen for the statistical processing of the data was Gries' (2009: 248–252) Hierarchical Configurational Frequency Analysis using the R software (R Core Team 2015).

Let us now move on to the accessibility of referents and how this influences pronoun omission. Table 3 sets out the distribution of omitted and overt pronouns per level of accessibility. The majority of cases of omission refer to highly accessible entities or events in both varieties, with only a few cases at the intermediate and low levels, and this distribution is statistically significant ($\chi^2(7) = 36.44, p < .001$). Overt pronouns behave similarly in this respect, since most cases occur at the highest level of accessibility, but omitted pronouns seem to be more strongly affected by the accessibility of referents. Indian English and Singapore English behave differently: there is a higher percentage of omitted pronouns at the highest level of accessibility and lower percentages at the intermediate and low levels in Singapore English than in Indian English.

Table 2: Distribution of omitted and overt pronouns per syntactic function.

FUNCTION	VARIETY			
	IndE		SgE	
	<i>Omitted</i>	<i>Overt</i>	<i>Omitted</i>	<i>Overt</i>
<i>Subject</i>	6.3% (161)	93.7% (2386)	9.2% (262)	90.8% (2578)
<i>Direct Object</i>	12.9% (44)	87.1% (296)	7.3% (24)	92.7% (306)
<i>Total</i>	7.1% (205)	92.9% (2682)	9% (286)	91% (2884)

Table 3: Distribution of omitted and overt pronouns per level of accessibility.

ACCESSIBILITY	VARIETY			
	IndE		SgE	
	<i>Omitted</i>	<i>Overt</i>	<i>Omitted</i>	<i>Overt</i>
<i>High</i>	7.4% (182)	92.6% (2273)	9.8% (269)	90.2% (2485)
<i>Intermediate</i>	6.9% (21)	93.1% (284)	5.7% (16)	94.3% (267)
<i>Low</i>	1.6% (2)	98.4% (125)	0.8% (1)	99.2% (132)

Focusing now on each of the accessibility factors individually, that is, the saliency of the referent in the discourse (Table 4) and the tightness of the link between the anaphoric element and its antecedent (Table 5), the latter seems to have a stronger influence on pronoun omission than the former. Most omitted pronouns refer to salient entities in both varieties, but in Indian English there is also a high percentage of omission (9.6%) with non-salient referents, and this distribution is statistically significant ($\chi^2(4) = 15.18, p < .01$). Therefore, it seems that saliency affects pronoun omission more strongly in Singapore English. Regarding the tightness of the link, it seems to have a strong influence in both varieties: the frequency of omitted pronouns increases to approximately 13% when the link between the anaphoric element and its antecedent is tight (as in example (11)), and it decreases to approximately 5% when the link is loose (as in (12)), and this is the case in both Singapore English and Indian English ($\chi^2(4) = 19.77, p < .001$).

- (11) *What are we looking at then you know we have this conflicting uhm aspirations in in women nowadays. You know they_i want to work and Ø_i want to have a child* (ICE-SIN:S1B-036#122–123:1:A)
- (12) G: *Bring her_i never mind lah Easier what Easier for you to handle*
 H: *Tomorrow uh*
 G: *Tomorrow bring to church uh*
 H: *Ya ya*
 G: *Oh you are uh uhm*
 H: *Uh uhm*
 G: *Ya finally I never see Ø_i* (ICE-SIN:S1A-035#308–317:3)

Table 4: Distribution of omitted and overt pronouns with respect to the saliency of the referent.

SALIENCY	VARIETY			
	IndE		SgE	
	Omitted	Overt	Omitted	Overt
<i>Salient</i>	7.2% (189)	92.8% (2419)	9.6% (278)	90.4% (2620)
<i>Non-salient</i>	9.6% (16)	90.4% (151)	4.4% (7)	95.6% (151)

Table 5: Distribution of omitted and overt pronouns with respect to the tightness of the link.

LINK	VARIETY			
	IndE		SgE	
	<i>Omitted</i>	<i>Overt</i>	<i>Omitted</i>	<i>Overt</i>
<i>Tight</i>	12.9% (133)	87.1% (900)	13.5% (133)	86.5% (854)
<i>Loose</i>	4.6% (8)	95.4% (165)	5.1% (9)	94.9% (166)

Finally, there is a context in which pronoun omission appears to be particularly frequent, and this is when the anaphoric element and its antecedent are located in coordinate clauses, as in (13):

- (13) *That uh <, > simply one fellow with a phone will hire some guys; <, > exactly <, > and send Ø; there for one year (ICE-IND:S1A-045#122:1:B)*

Table 6 displays the distribution of omitted and overt pronouns with respect to the syntactic relation between the anaphor and its antecedent. It can be seen that there is a higher percentage of omission in coordinate clauses in both varieties (44.6% in Indian English and 51.8% in Singapore English) and a lower one in embedding contexts and when there is not a syntactic relation between the anaphor and the antecedent ($\chi^2(7) = 494.37, p < .001$).

Table 6: Distribution of omitted and overt pronouns with respect to syntactic relation.

SYNTACTIC RELATION	VARIETY			
	IndE		SgE	
	<i>Omitted</i>	<i>Overt</i>	<i>Omitted</i>	<i>Overt</i>
<i>Embedding</i>	2.3% (4)	97.7% (168)	2.1% (4)	97.9% (185)
<i>Coordination</i>	44.6% (78)	55.4% (97)	51.8% (85)	48.2% (79)
<i>No Relation</i>	7.1% (59)	92.9% (772)	6.4% (51)	93.6% (743)

7. Discussion

The results presented in the previous section suggest that pronoun omission is not a widespread phenomenon in either Indian English or Singapore English: the frequency of omitted pronouns reaches 7.1% in the former variety and 9% in the latter. This means that, in most cases, speakers chose to use overt pronouns to refer to entities or events, even though pronoun omission results in more efficient and easier-to-process structures according to Hawkins' (2004) metric. Omitted pronouns are basically restricted to those occasions in which their referents can be easily recovered because of their accessibility in discourse: out of the 205 and 286 instances of omitted pronouns in Indian English and Singapore English, only 23 (11.2%) and 17 (5.9%) instances respectively do not refer to highly accessible entities. Overt pronouns behave in a similar fashion, that is, they are mostly used to refer to highly accessible antecedents, and they actually seem to be the unmarked option. There is, however, a difference between these two anaphoric devices: omitted pronouns are more strongly influenced by accessibility considerations than overt ones because the hearer needs additional help to coindex an omitted pronoun with its correct referent.

If we consider each of the accessibility factors individually, the saliency of the referent in discourse seems to play a less important role than the tightness of the link between the anaphoric element and its antecedent, and especially if the syntactic relation that links both elements is coordination. When the anaphor and the antecedent are located in coordinate clauses, pronouns are omitted in half the cases in both Indian English and Singapore English. This is not surprising if we take into account that many accessibility factors apply at the same time in coordination contexts: in most cases the antecedent is in subject position (although it can also be in other positions, as in example (13) above) and in the immediately preceding clause (except when one or both of the coordinate clauses contain embedded clauses within them). Moreover, coordinate clauses are usually connected by a conjunction, which increases their pragmatic cohesion. All of these factors, together with the fact that pronouns can also be omitted in subject position in

coordination in Standard English, make coordinate clauses a particularly favourable context for omitted pronouns.

Indian English and Singapore English display a similar distribution of omitted and overt pronouns, although there are some differences between them. First, there is a higher total percentage of omitted pronouns in Singapore English than in Indian English. Second, pronoun omission is more strongly affected by accessibility considerations in Singapore English than in Indian English: omitted pronouns refer more frequently to highly accessible entities in the former than in the latter variety. These facts point to the conclusion that pronoun omission is a more regular feature in Singapore English than in Indian English, and this may be related to the different sociolinguistic circumstances of the varieties. In India, English is mostly used as a second language and is restricted to specific purposes and domains of life such as government and administration, politics, education, etc. Indian English is influenced by conservative attitudes that result in that the innovative features of its grammar are not always accepted as correct (Mukherjee 2007). In Singapore, however, English is increasingly being acquired as a first language and it is also used in more intimate and informal domains of life. The variety of English spoken in Singapore, with all its characteristic features, carries positive connotations as a marker of the nation's cultural identity, and the innovations that are part of its grammar are more frequently accepted as correct (Schneider 2007: 156). The conservative attitudes towards English in India may have an inhibitory effect on the features that differentiate it from other varieties, while in Singapore such restrictions do not generally apply. This difference between the varieties may account for the fact that pronoun omission is more stable and more frequent in Singapore English than in Indian English.

Finally, section 2 introduced Hawkins' (2004) Performance Grammar Correspondence Hypothesis, which proposed that efficient linguistic features may end up being conventionalized in the grammar. Pronoun omission is such a feature when the referents of omitted pronouns are accessible and, therefore, it was hypothesized that omitted pronouns could be conventionalized as an anaphoric device referring to highly accessible entities in the grammars of Indian English and Singapore English. This hypothesis seems not to be supported by the data because

pronoun omission is not a very frequent phenomenon in either variety and overt pronouns are actually the unmarked option to refer to this class of antecedents. There is one exception, however, and this is when the omitted pronoun and its antecedent are in coordinate clauses: in this context pronouns are omitted in half the cases in both Indian English and Singapore English. This suggests that pronoun omission may be a conventionalized option in coordinate clauses, which provides support to Hawkins' (2004) hypothesis.

8. Conclusions

The aim of this paper was to analyze the distribution of omitted and overt pronouns with respect to the accessibility of their referents and to ascertain whether pronoun omission is a conventionalized anaphoric device to refer to a certain class of antecedents in Indian English and/or Singapore English, in accordance with Hawkins' (2004) Performance Grammar Correspondence Hypothesis. The results showed that omitted pronouns were coindexed almost exclusively with highly accessible antecedents in discourse but that overt pronouns were still the unmarked anaphoric elements used for this class of entities. In coordinate clauses, however, pronouns were omitted in half the cases. This suggests that pronoun omission may be a conventionalized option in this context. Indian English and Singapore English behaved similarly in this respect, but there were some differences between them: pronoun omission was more frequent and omitted pronouns were used more consistently to refer to highly accessible entities in Singapore English than in Indian English. These differences suggest that pronoun omission is a more established feature in Singapore English, a fact that may be related to the distinct sociolinguistic conditions in India and Singapore.

The present study has provided data on the distribution of omitted pronouns in Indian English and Singapore English, but there are still many unanswered questions. Accessibility considerations clearly affect pronoun omission in the two varieties of English studied, but other factors, such as agreement or frequency, could also have an impact.

In addition, the results presented here should be compared with real corpus data from native varieties of English in order to ascertain whether they in fact differ from high-contact varieties. Williams (1988) compared the speech of native speakers of English with that of second language and Singapore English speakers, and she found differences in frequency and in the contexts in which pronouns were omitted. Her study, however, was restricted to third person pronouns in subject position, so further research is still needed with respect to this and other issues.

Databases and tools

International Corpus of English – the Indian Component 2002. Project coordinated by Prof. S. V. Shastri at Shivaji University, India, and Prof. Dr. Gerhard Leitner at Freie Universität Berlin, Germany. Available online at <<http://ice-corpora.net/ice/download.htm>>.

International Corpus of English – the Singaporean Component 2002. Project Coordinated by Prof. Paroo Nihilani, Dr. Ni Yibin, Dr. Anne Pakir and Dr. Vincent Ooi at The National University of Singapore, Singapore. Available online at <<http://ice-corpora.net/ice/download.htm>>.

Kortmann, Bernd / Lunkenheimer, Kerstin (eds) 2013. *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology <<http://www.ewave-atlas.org>> (accessed October 20, 2014).

R Core Team 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <<http://www.R-project.org>> (accessed June 22, 2015).

Scott, Mike 2012. *WordSmith Tools*. Version 6. Liverpool: Lexical Analysis Software.

References

- Ariel, Mira 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Ariel, Mira 1994. Interpreting Anaphoric Expressions: A Cognitive versus a Pragmatic Approach. *Journal of Linguistics*. 30/1, 3–42.
- Ariel, Mira 2001. Accessibility Theory: An Overview. In Sanders, Ted / Schliperoord, Joost / Spooren, Wilbert (eds) *Text Representation*. Amsterdam: John Benjamins, 29–87.
- Bao, Zhiming 2001. The Origins of Empty Categories in Singapore English. *Journal of Pidgin and Creole Languages*. 16/2, 275–319.
- Clark, Herbert H. / Sengul, C. J. 1979. In Search for Referents for Nouns and Pronouns. *Memory and Cognition*. 7/1, 35–41.
- Cole, Melvyn D. 2010. Thematic Null Subjects and Accessibility. *Studia Linguistica*. 64/3, 271–320.
- Dahl, Östen 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam: John Benjamins.
- Talmy Givón (ed.) 1983. Topic Continuity in Discourse: An Introduction. In Givón, Talmy (ed) *Topic Continuity in Discourse: A Quantitative Cross-language Study*. Amsterdam: Benjamins, 5–41.
- Gries, Stephan T. 2009. *Statistics for Linguistics with R: A Practical Introduction*. Berlin: Mouton de Gruyter.
- Gundel, Jeanette K. / Tarone, Elaine 1983. Language Transfer and the Acquisition of Pronominal Anaphora. In Gass, Susan / Selinker, Larry (eds) *Language Transfer and Language Learning*. Rowley, MA: Newbury House, 281–296.
- Haegeman, Liliane 1994. *Introduction to Government and Binding Theory* (2nd ed). Oxford: Blackwell.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, John A. 2009. An Efficiency Theory of Complexity and Related Phenomena. In Sampson, Geoffrey / Gil, David / Trudgill, Peter (eds) *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 252–268.

- Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
- Huang, C.-T. James 1984. On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry*. 15/4, 531–574.
- Huddleston, Rodney / Pullum, Geoffrey K. et al. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Klein, Wolfgang / Perdue, Clive 1997. The Basic Variety (or: Couldn't Natural Languages Be Much Simpler?). *Second Language Research*. 13/4, 301–347.
- Lange, Claudia 2012. *The Syntax of Spoken English Indian English*. Amsterdam: John Benjamins.
- Leimgruber, Jakob 2013. *Singapore English: Structure, Variation, and Usage*. Cambridge: Cambridge University Press.
- Miestamo, Matti 2008. Grammatical Complexity in a Cross-linguistic Perspective. In Miestamo, Matti / Sinnemäki, Kaius / Karlsson, Fred (eds) *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 23–41.
- Miestamo, Matti / Sinnemäki, Kaius / Karlsson, Fred (eds) 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins.
- Mukherjee, Joybrato 2007. Steady States in the Evolution of New Englishes: Present-Day Indian English as an Equilibrium. *Journal of English Linguistics*. 35/2, 157–187.
- Ruppenhofer, Josef / Michaelis, Laura A. 2010. A Constructional Account of Genre-Based Argument Omissions. *Constructions and Frames*. 2/2, 158–184.
- Sampson, Geoffrey 2009. A Linguistic Axiom Challenged. In Sampson, Geoffrey / Gil, David / Trudgill, Peter (eds) *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 243–251.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 2013. English as a Contact Language: The “New Englishes”. In Schreier, Daniel / Hundt, Marianne (eds) *English as a Contact Language*. Cambridge: Cambridge University Press, 131–148.

- Toole, Janine 1996. The Effect of Genre on Referential Choice. In Fretheim, Thornstein / Gundel, Jeanette (eds) *Reference and Referent Accessibility*. Amsterdam: Benjamins, 263–290.
- Trudgill, Peter 2009a. Sociolinguistic Typology and Complexification. In Sampson, Geoffrey / Gil, David / Trudgill, Peter (eds) *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 98–109.
- Trudgill, Peter 2009b. Vernacular Universals and the Sociolinguistic Typology of English Dialects. In Filppula, Markku / Klemola, Juhani / Paulasto, Heli (eds) *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London: Routledge, 304–322.
- Williams, Jessica 1987. Non-native Varieties of English: A Special Case of Language Acquisition. *English World-Wide*. 8/2, 161–199.
- Williams, Jessica 1988. Zero Anaphora in Second Language Acquisition: A Comparison among Three Varieties of English. *Studies in Second Languages Acquisition*. 10/3, 339–370.
- Xu, Liejiong 1986. Free Empty Category. *Linguistic Inquiry*. 17/1, 75–93.