

Pronoun omission in high-contact varieties of English

Complexity versus efficiency

Iván Tamaredo

University of Santiago de Compostela

This paper considers pronoun omission in different varieties of English. It argues that omitted pronouns simplify structures if their referents are accessible in discourse, which explains the greater frequency of this grammatical feature in high-contact varieties of English, spoken in speech communities with a history of high numbers of second-language users. A corpus study of two high-contact varieties, Indian English and Singapore English, and a low-contact one, British English, is conducted in order to examine the distribution of omitted and overt pronouns. As expected, pronoun omission is more frequent in the high-contact varieties than in British English. Moreover, pronouns are omitted almost exclusively when they have highly accessible referents as antecedents, which is not a conventionalized feature of the grammars of Indian or Singapore English, where overt pronouns are the default choice when referring to antecedents.

Keywords: pronoun omission, complexity, efficiency, contact, varieties of English, Indian English, Singapore English, British English

1. Introduction

The present article addresses pronoun omission in different varieties of English from the perspective of its complexity and communicative efficiency. It will be argued that omitted pronouns make linguistic structures simpler and more efficient if referents can be recovered from their linguistic and/or extra-linguistic context. On the basis of prior studies which maintain that language contact results in grammatical simplification, it will be shown by means of a survey based on *The Electronic World Atlas of Varieties of English* (eWAVE) that pronoun omission is more frequently attested in high-contact than in low-contact varieties of English, thus supporting earlier claims that high-contact varieties tend to be simpler than

low-contact ones. In addition, two high-contact varieties, Indian English (IndE) and Singapore English (SgE), and a low-contact one, British English (BrE), will be explored in depth to provide quantitative data on the distribution of omitted pronouns. The results of this study will also reflect simplification effects of language contact: pronoun omission is more frequent in the high-contact varieties than in BrE. Moreover, the referents of omitted pronouns are almost exclusively highly accessible ones, but overt pronouns are still the default option. The one exception is coordination, where omission may be conventionalized as part of the grammars of the varieties under study.

This paper is structured as follows. Section 2 will describe the feature of pronoun omission as it is understood in the present study, and it will show examples of relevant cases. Section 3 will give a brief overview of the concepts of linguistic complexity and communicative efficiency, which will then be applied to the analysis of the feature. Section 4 will examine the relation between contact, complexity, and efficiency by means of a survey of pronoun omission in varieties of English. Sections 5, 6, and 7 will deal with the methodology, results and discussion of a corpus study on the distribution of omitted and overt pronouns in SgE, IndE and BrE. Finally, Section 8 will provide some concluding remarks.

2. Pronoun omission

For the present purposes, the term “pronoun omission” will be applied to situations where there is a gap in a structure that could be filled by an overt personal pronoun. The following types of pronouns will be distinguished on the basis of their function and referential properties: referential pronouns in subject position (1), non-referential pronouns in subject position (2), referential pronouns in direct object position (3), and non-referential pronouns in direct object position (4)¹.

- (1) <ICE-IND:S1B-055#11:1:C> And another thing I would like to know [...] whether they are able to plan any <,> new systems <,> that will enable to increase the fertility <,> in order to increase [the requirements of the crude <,> in this country]₁ <,>
 <ICE-IND:S1B-055#12:1:C> The honourable minister has made a statement saying within three years Ø₁ have increased to <,> up to sixty percent of the last <,> pre previous production <,>

1. The examples are taken from the Indian and Singaporean components of the *International Corpus of English* (ICE), unless otherwise stated. See Section 5 for further detail about the corpus used.

- (2) <ICE-SIN:S1A-095#183:1:B> Then uhm after that she's free to move about and this time it'll be quite different in the sense that she won't have those uhm like the angle equipments attached so she'll only have like uh six electrodes attached to her and that's it
 < ICE-SIN:S1A-095#184:1:A> Ya
 < ICE-SIN:S1A-095#185:1:B> So probably \emptyset will be [...] on made easier for her to move around
- (3) <ICE-IND:S1A-045#122:1:B> That uh <,> simply one fellow with a phone will hire **some guys**_i <,> exactly <,> and send \emptyset _i there for one year
- (4) <ICE-SIN:S1B-055#10:1:B> The teachers are given very wide syllabus to cover and they have to go very quickly through them and of course the average uh pupils uh would find it very difficult to cope up²

All instances of omission in object position are considered here, but in subject position only instances in finite clauses are taken into account.³

Pronoun omission has been widely analyzed cross-linguistically, mainly with focus on syntactic restrictions (cf. Chomsky 1981; Huang 1984; Rizzi 1986; Jaeggli and Safir 1989; Neeleman and Szendrői 2007), but also on more pragmatic and/or cognitive factors (cf. Huang 1992, 2000; Cole 2009, 2010; Ariel 1988, 1990, 1994, 2001). Generative accounts have highlighted the connection between pronoun omission and agreement, particularly in languages with rich agreement between the verb and its arguments. This is not always so, however. Languages such as Mandarin Chinese allow omission both in subject and object positions even though they do not have agreement (Huang 1984). In these languages, verbal morphology does not play a role, and despite syntactic restrictions, whether the pronoun is omitted or not is mostly ruled by issues of a pragmatic nature. Additionally, Cole (2010: 280) points out that even in languages with rich agreement, this is not always sufficient to recover the referent of an omitted pronoun due to (a) ambiguity (that is, more than one potential antecedent with the same person and number features) or (b) syncretism (that is, when the same form of the suffix is used for different feature combinations: e.g. the form *comía* in Spanish [*I/he/she/it ate*] can be used with a first person singular and a third person singular subject). In these cases the referent must be retrieved from context. Therefore, the main difference between languages that allow pronoun omission and those that do not is whether they rely on context for the identification of referents (when there is no verb-morphological information).

2. In Example (4) there is an overt pronoun since no instances of non-referential omitted pronouns were found in direct object position in the corpus.

3. Instances of subject (referential) pronoun omission in non-finite clauses and in imperatives are excluded from the present study.

Ariel (1988, 1990, 1994, 2001) has developed an Accessibility theory of the availability of referents in discourse. Its central tenet is that all referential expressions guide the addressee to the correct interpretation by indicating the level of accessibility of a particular referent in his or her memory. The main factors are referent saliency and the tightness of the link between the anaphoric element and its antecedent. Saliency is increased if the entity referred to is a discourse topic, if it is animate (especially human) or in subject position, and if there are no other potential antecedents in the surrounding context, among other factors. With respect to the connection between the anaphoric element and its antecedent, it is tighter if they are in syntactically connected clauses (through subordination or coordination), in sentences connected by pragmatic means, and if there is a short distance between them. Ariel (2001: 31) establishes the following accessibility marking scale:

Full name + modifier > full name > long definite description > short definite description > last name > first name > distal demonstrative + modifier > proximate demonstrative + modifier > distal demonstrative + NP > proximate demonstrative + NP > distal demonstrative (-NP) > proximate demonstrative (-NP) > stressed pronoun + gesture > stressed pronoun > unstressed pronouns > cliticized pronoun > verbal person inflections > zero

The expressions at the beginning of the scale, such as full name + modifier, are used to refer to referents with low accessibility, whereas those at the end refer to highly accessible ones. Zero is the expression that includes pronoun omission, which marks the highest level of accessibility.

Present-day English is considered a non-pro-drop language, that is, one in which omitted pronouns are not allowed by the grammar. However, they occur in certain contexts. Instances of pronoun omission can be found in casual style in subject position at the beginning of main clauses, both declarative and interrogative (Huddleston and Pullum et al. 2002: 1540–1541). It is the default option in imperatives, and it also occurs in non-finite clauses in all styles. Finally, a pronoun can be omitted in coordinate clauses, but only if it is the subject of the second conjunct and if it is coreferential with that of the first one.

3. Linguistic complexity

The concept of complexity has played an important role in linguistic research. One central assumption of 20th-century linguistics was that all languages are equal with respect to their overall grammatical complexity. This assumption was based on the idea that if a language is more complex than others with respect to one subsystem of the grammar (for instance, morphology), then it must have a higher degree of

complexity in some other subsystem (for instance, syntax). Around the turn of the 21st century, a number of studies challenged the assumption of complexity invariance and demonstrated that languages do differ in the complexity of specific grammatical subsystems, but that there are not necessarily trade-offs between subsystems, that is, a language does not necessarily compensate for a simpler morphology with a more complex syntax (cf. Fenk-Oczlon and Fenk 2008; Sinnemäki 2008; Dahl 2009; Nichols 2009).

In order to make meaningful comparisons between languages with respect to their grammatical complexity, various metrics of complexity have been put forward in the literature. Even though they do not operationalize complexity in the same manner, many of them include similar principles.

3.1 Metrics of grammatical complexity: What do simple and complex mean?

Complexity is not an easy concept to define and measure. Despite the recent proliferation of metrics in the literature on grammatical complexity, there are two general types of principles that subsume many of the criteria postulated so far. These can be termed principles of “economy” and principles of “transparency”.

Economy (for this label see, among others, Kusters 2003: 22–25) can be defined in terms of the amount of obligatory coding of grammatical categories and relations in a language. From this point of view, simplicity increases if fewer semantic or syntactic categories such as tense, person, case, etc., are distinguished in a grammar, and if there are fewer distinctions within each category (for example, a language that distinguishes between nominative and accusative case is simpler than one that distinguishes between nominative, accusative, and dative). On the other hand, a language is more complex if it gives overt marking to categories that other languages leave unmarked. Economy figures prominently in several metrics of complexity, such as in Kusters’ (2003) Economy principle, Miestamo’s (2008) Fewer Distinctions principle, McWhorter’s (2007) Overspecification factor, Szmrecsanyi and Kortmann’s (2009) Grammaticity index, and Trudgill’s (2009a, 2011) conceptualization of the appearance of new morphological categories as a complexifying process, among others.

Transparency (Kusters 2003: 26–30) refers to the relation between meanings and their formal marking. Simplicity increases when one grammatical meaning is coded by one form, and it decreases when there is no one-to-one relation, for instance in cases of fusion, fission, homonymy, and allomorphy (Kusters 2003: 26–30). Irregular marking also increases complexity in that it implies the existence of extra idiosyncratic forms for a grammatical category that is already coded by a regular, consistent pattern. Transparency is an integral component of Kusters’

(2003) Transparency principle, Miestamo's (2008) One-Meaning – One-Form principle, McWhorter's (2007) Irregularity factor, Szmezsanyi and Kortmann's (2009) Transparency index, and Trudgill's (2009a, 2011) complexifying processes of irregularization, and the increase of syntagmatic redundancy and morphological opacity, among others.

Economy and transparency are two of the criteria that form the basis of Hawkins' (2004, 2009) metric of communicative efficiency. This metric will be used here to analyze the feature of pronoun omission because it allows departures from an ideal level of transparency in certain contexts, which is necessary in order to understand the occurrence of omitted pronouns.

3.2 Hawkins' metric of communicative efficiency

Hawkins' metric of communicative efficiency is grounded in processing and cognitive motivations.⁴ The concept of communicative efficiency is key. Communication is considered to be efficient when the message conveyed by the speaker is transmitted to the addressee with structures that are easy to process, allowing fast information transmission (Hawkins 2009: 253). This definition seems to imply that efficiency always involves minimizing the formal complexity of structures so that they are easier for the speaker to articulate. However, for communication to be efficient, it must also be successful, that is, the addressee must be able to decode information. In some cases, this means that structures must contain more forms to guide the addressee to the correct interpretation. Efficiency and complexity are, therefore, two separate concepts that must not be confused. A structure is more complex if it has more forms (phonemes, morphemes, phrases, etc.) with their associated meanings, and if the domains for the identification of the syntactic and semantic connections between these forms are larger. This is so because a structure with more forms and larger domains is more costly to process. Efficiency may involve more or less complexity, depending on the message that the speaker tries to convey and the minimum formal complexity that is needed for the successful transmission of this message (Hawkins 2004: 9).

Hawkins' metric contains three principles of efficiency, formulated in terms of processing preferences: Minimize Domains, Maximize On-line Processing, and Minimize Forms. In the present study, only the Minimize Forms principle is relevant. Hawkins defines it as follows:

4. What follows is a summary of the main points of Hawkins' metric, and especially of those that are relevant for the present study. For further details, see the entries for Hawkins in the bibliography.

The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. 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)

This is a principle of economy that includes another dimension not considered in most metrics of complexity: besides the preference for fewer “forms with unique conventionalized property assignments”, that is, fewer grammatical distinctions, the Minimize Forms principle also favors shorter forms with fewer phonemes, morphemes, words, or phrases. The rationale behind this principle is that processing linguistic forms and their meanings requires some effort. By decreasing the number of forms in a structure (and in the grammar as a whole), the speaker needs less effort and less time for articulation. However, for a minimal structure to be efficient, the addressee must also be able to decode the message that it conveys, that is, there is a limit to form minimization. In order to ease the task of decoding a minimal structure, the addressee can resort to other sources of information, such as the linguistic and extra-linguistic contexts, and fill in the gaps left by the speaker. Hawkins (2004: 41) calls this inferential procedure “processing enrichment”, a notion that must be taken into account when trying to understand why and when structures can be minimized. There are several factors that facilitate the ease with which processing enrichments can be effected, the most pertinent to our current study being high entity or event accessibility in the discourse.⁵

Pronoun omission results in structures with fewer forms: there would be an extra word in Examples (1)–(3) above if pronouns were overt. In principle, this is a departure from ideal transparency as defined above, but Hawkins’ metric of communicative efficiency allows for such departures as long as the structure can be enriched using the information already active in the context. According to Hawkins’ Minimize Forms principle, structures with fewer forms are easier to process because they are less costly for the speaker to articulate. However, for communication to be successful, the referents of pronouns in positions such as subject and direct object must be identified. When pronouns are omitted, the addressee must resort to the context for the identification of their referents and, if they are accessible, the structure can be enriched in processing and the addressee can decode the message. Therefore, pronoun omission results in simpler structures that are also efficient if the referents of omitted pronouns are accessible in the discourse.

5. See Section 2.

Finally, an important part of Hawkins' theory is what he terms the "Performance Grammar Correspondence Hypothesis", which proposes that grammars are fixed conventions, reflecting the processing preferences of users: speakers prefer to produce structures that are communicatively efficient, and these efficient structures then become part of the grammar (Hawkins 2004: 3). One of the goals of the present article is to test this hypothesis with respect to pronoun omission in two high-contact varieties of English, IndE and SgE.

3.3 Complexity, efficiency, and contact

Several studies have demonstrated that languages can differ with respect to their grammatical complexity, and that complexity can also vary in the history of a language. The challenge is to identify the factors that cause differences in complexity. Several phenomena have been shown to have an effect,⁶ but the one that has probably received most attention is language contact (cf., for example, the collection of papers in Miestamo, Sinnemäki, and Karlsson 2008). A history of contact in a speech community causes the grammar of the language spoken in that community to become simpler "due to the relative inability of adult humans to learn new languages perfectly" (Trudgill 2009a: 99). It is important to note, however, that different types of contact bring about different outcomes. Trudgill (2011: 34) proposes a typology of contact situations which affect grammars in different ways. Simplification occurs in high-contact situations where non-native language learning and use by adults is widespread. This is due to the fact that the language learning abilities of adult speakers are less than perfect, and therefore they tend to disregard complex features. In contrast, long-term stable contact involving child bilingualism may cause complexification because of the addition of new categories due to the borrowing of features from one language to another, or due to contact-induced grammaticalization, a process whereby a new grammatical category emerges in a language on the model of another (Heine and Kuteva 2005: 7). Finally, grammatical complexity may increase spontaneously in languages spoken in low-contact communities, where there is no influence of another language (Trudgill 2011: 64).

So far we have dealt with the influence of contact on grammatical complexity but not with the contact/communicative efficiency connection. The concept of efficiency can be understood in terms of a trade-off between processing cost and communicative benefits: if two structures codify the same information, the one that requires the least effort on the part of users is the most efficient. This means that, if this structure is used, the speaker will be able to transmit the same message to the addressee with fewer processing resources. Choosing the most efficient

6. For a review of different factors that influence grammatical complexity, see Sampson (2009).

structures allowed by the grammar of a language is beneficial for speakers (the basis of Hawkins' Performance Grammar Correspondence Hypothesis).⁷ High-contact situations are characterized by a high number of second-language users. In order to communicate, these users have at their disposal the same cognitive machinery as native speakers, but they have to deal with the additional difficulty of speaking in a language that is not their mother tongue. It would not be surprising, therefore, if these speakers selected grammatical features that result in efficient structures more frequently than native speakers, because these features allow them to compensate to a certain degree for the added effort of using a second language. Note in this regard Williams (1987, 1988), who argues that two complementary production and comprehension principles, economy and hyperclarity, guide communication. The former makes "production more efficient and economical" (Williams 1987: 169), while the latter keeps it transparent and decodable for the addressee. In spontaneous speech, where the time to plan and transmit the message is limited, speakers (even native ones) tend to give priority to those structures that allow them to maximize economy without losing clarity and, therefore, save processing resources that might be necessary for some other task. Non-native speakers have to face the same production pressures, but in addition they are also "constrained by limited proficiency" (Williams 1988: 365). The result is that they compensate for this double effort by producing economical, but still decodable, structures more often than native speakers, and these efficient structures in the speech of second-language speakers may in turn become conventionalized in the grammars of L2 varieties of English.

4. Pronoun omission in varieties of English

The study of varieties of English is a promising field of research in order to understand the connection between contact and complexity/efficiency. The spread of English around the world has caused the emergence of new varieties of English in the former British colonies, often as a result of different contact types. Given that this has been a relatively recent development, we have information regarding the sociohistorical conditions in which these varieties developed, and this in turn means that we can establish connections between social determinants and their putative effect on grammatical complexity. These can be extrapolated to other languages for which sociolinguistic data are not available.

If we hypothesize that pronoun omission is a grammatical feature that results in simpler and efficient structures,⁸ we first need to determine whether it is frequent in

7. See Section 3.2.

8. See Section 3.2.

the speech of second-language speakers. Williams (1988) compared the speech of native speakers of English, second-language speakers of English with different first languages, and speakers of SgE, a high-contact variety. She studied the distribution of omitted third person pronouns in subject position and found that they were less frequent in native than in second-language and SgE speakers. Moreover, the former group omitted pronouns in a more restricted set of contexts, mostly in coordinate clauses with an antecedent in the preceding clause. Gundel, Stenson, and Tarone (1984) studied the omission of direct object pronouns by second-language speakers of English with Spanish as a first language (Gundel and Tarone 1983) and children with English as a first language learning French as a second language (Gundel, Stenson, and Tarone. 1984). They found that both groups omitted pronouns in direct object contexts when they spoke their second language. Finally, Zyzik (2008) also focused on the omission of direct object pronouns in second-language speakers of Spanish with English as a first language. She also found that they omitted direct object pronouns, especially when their proficiency in the second language was not high. These studies found that pronoun omission, both in subject and direct object position, was characteristic of the speech of second-language users. More importantly, this is the case irrespective of the specific L1/L2 combination, a fact that suggests that transfer from the first to the second language may not be a decisive factor and that a higher frequency of pronoun omission seems to be a consequence of second language acquisition and use.⁹

Accordingly, I suggest that it should be more frequent in varieties of English with a history of contact also, that is, with high numbers of non-native users. Using eWAVE as a source of data, I conducted a survey of pronoun omission in varieties of English, including both subject and direct object pronouns. The varieties were categorized either as low- or high-contact on the basis of Trudgill's (2009b) classification, but with one minor difference. As high-contact varieties Trudgill includes indigenized L2 varieties, shift varieties,¹⁰ dialect-contact varieties,¹¹ and pidgins

9. The role of transfer, however, must also be recognized. Williams (1988: 349) mentions that the second-language speakers of English that she studied differ in the frequency with which they dropped pronouns, and that this "may well be the result of L1 influence". The substrate languages of IndE and SgE, the high-contact varieties that are the focus of the present paper, may have facilitated the emergence of pronoun omission in these varieties (cf. Moag and Poletto 1991; Bao 2001)

10. Those varieties that were L2 varieties in the past, but that since then have been adopted as a first language by the majority of the speech community.

11. Those varieties that are the result of the convergence of different dialects of English in the same speech community.

and creoles; as low-contact varieties, he includes traditional L1 dialects.¹² Since pronoun omission was a characteristic of second-language users, dialect-contact varieties were included in the low-contact category with traditional L1s, because these two types of varieties were not affected by the process of non-native language learning. Table 1 shows the frequency of attestation of pronoun omission in the two categories of varieties of English. Less than 40 percent of low-contact varieties attest pronoun omission, while more than 70 percent of high-contact varieties do so, and these results are statistically significant ($\chi^2(1) = 6.74, p < .01$).¹³ The fact that pronoun omission is more frequently attested in high- than in low-contact varieties is in line with what we discussed above: pronoun omission is a grammatical feature that simplifies structures and increases their efficiency, provided that the referent of the pronoun is accessible, and second-language speakers of English (and other languages) omit pronouns more frequently than native speakers.

Table 1. Attestation of pronoun omission in varieties of English

Varieties	Pronoun omission	
	Attested	Not attested
Low-contact	38.9% (7)	61.1% (11)
Traditional L1s	40% (4)	60% (6)
Dialect-contact L1s	37.5% (3)	62.5% (5)
High-contact	72.4% (42)	27.6% (16)
Shift L1s	85.7% (12)	14.3% (2)
Indigenized L2s	72.2% (13)	27.8% (5)
Pidgins	71.4% (5)	28.6% (2)
Creoles	63.2% (12)	36.8% (7)

The rest of this paper is devoted to a corpus study of the distribution of omitted pronouns in SgE, IndE, and BrE. The research questions are:

12. Varieties spoken in isolated communities, mainly in certain areas of the British Isles or North America.

13. The test used for the statistical processing of all the data in the present study was the chi-squared test.

- Is there an observable simplifying effect of language contact, so that pronoun omission is more frequent in the high-contact varieties (SgE and IndE) than in BrE?
- Do speakers omit pronouns efficiently, that is, when their referents can be recovered from the linguistic or extra-linguistic context?
- If the answer to the previous question is positive, are omitted pronouns a conventionalized referential device of the grammars of SgE and IndE to refer to accessible antecedents? In Section 3.2, Hawkins' Performance Grammar Correspondence Hypothesis was presented. It suggested that the structures preferred by users end up becoming part of the grammar of languages. If this is the case, high-contact varieties of English may have conventionalized pronoun omission as part of their grammars due to the fact that second-language speakers tend to omit pronouns.

5. Materials and methodology

All the instances of omitted and overt pronouns found in a set of texts taken from the Singaporean (ICE-SIN), the Indian (ICE-IND), and the British (ICE-GB) components of the *International Corpus of English* (ICE) were analyzed, and the results were compared in order to determine whether there were different patterns of use. IndE and SgE were chosen because, according to eWAVE, both present a high frequency of pronoun omission, and, even though they are both high-contact varieties, they belong to different subtypes: IndE is an indigenized L2 and SgE is in the process of becoming a shift L1 variety. This provided an opportunity to observe whether and to what extent they differ with respect to the frequency and conditions of use of omitted pronouns. Additionally, BrE was included in the study in order to (1) compare high- and low-contact varieties of English with respect to the distribution of omitted and overt pronouns, and (2) to examine whether there are differences due to the simplifying effect of language contact. BrE is a native variety and, according to Trudgill's (2009b) classification, a subtype of dialect-contact, which for the purposes of the present study is included in the low-contact category.¹⁴

The concordance software *Wordsmith Tools* (Scott 2012) was used in order to retrieve instances of overt pronouns, whereas instances of omitted pronouns were identified manually. The corpus analyzed comprised sixty texts (approximately a 130,000 words) taken from ICE-SIN, ICE-IND, and ICE-GB. Each ICE component

14. See Section 4.

is composed of one million words of spoken and written language and attempts to be representative of one variety of English. All components follow the same design and code for grammatical annotation. Of the sixty texts selected, twenty belonged to ICE-SIN, twenty to ICE-IND, and twenty to ICE-GB. These texts were chosen evenly with respect to register and mode, so that, although small, the corpus was balanced. For each variety, five texts each were spoken and informal, spoken and formal, written and informal, and 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, spoken formal texts from S1B (mainly from parliamentary debates and legal cross-examinations), written informal texts from W1B (from the genre of social letters) and, finally, written formal texts from W2A, W2C, W2D, and W2E.

The accessibility of the referents of omitted and overt pronouns was measured following Accessibility theory,¹⁵ which for our purposes was operationalized in the following terms. The accessibility of an entity or event increases if it is salient in the current discourse and if there is a tight link between the anaphoric element and its antecedent. The saliency of the referent is increased if:

1. It is a discourse topic: an entity was considered to have high topicality if it was mentioned at least twice in the four clauses preceding the occurrence of the anaphoric element (Example 5).
2. It is in subject position (Example 6).
3. It is animate rather than inanimate (Example 6).
4. It is emphasized by being in a syntactic construction that sets it apart from the other constituents of the clause, for example, by being preposed, postposed, by being part of a cleft or a pseudo-cleft sentence, etc. (Example 7).
5. 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 (Example 7).
6. It is coded using a low accessibility marker, for example, a full noun phrase rather than a pronoun (Example 7).
 - (5) <ICE-SIN:S1A-062#3:1: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
<ICE-SIN:S1A-062#4:1:A> So so already \emptyset_i told me said that he's going to be in San Francisco [...]
 - (6) <ICE-IND:S1A-010#51:1:C> And daily now **we_i** <.> le </.> read linguistic <,> \emptyset_i talk linguistic <,> \emptyset_i walk linguistic

15. See Section 2.

- (7) <ICE-SIN:W1B-003#41:1> I'll pray for your well-being ie safety.
 <ICE-SIN:W1B-003#42:1> Lionel almost got robbed.
 <ICE-SIN:W1B-003#43:1> Did he tell you ?
 <ICE-SIN:W1B-003#44:1> Drunk natives, \emptyset_i must have been towering way
 above him.

The tightness of the link between the anaphoric element and its antecedent is increased if:

7. The anaphor and the antecedent are in embedded clauses (Example 8).
8. They are in coordinate clauses (Example 9).
9. The sentences containing them are pragmatically cohesive, that is, if they are connected by conjunctive adverbs (Example 10) or if the antecedent is in an initiating speech act, such as a question, and the anaphor in a responding speech act, such as the answer to the question (Example 11).
10. There is a short distance in clauses between them, not more than four clauses (Example 11).
 - (8) <ICE-IND:S1A-099#268:2:C> [...] **they**_i wanted me to join \emptyset_i <,> before my exams.
 - (9) <ICE-IND:S1A-095#160:3:B> Yesterday also **she**_i was at their place and \emptyset_i came
 - (10) <ICE-IND:S1A-099#110:2:A> Yeah one of them writes <,> letters to the newspaper
 <ICE-IND:S1A-099#111:2:C> <indig> Accha </indig>
 <ICE-IND:S1A-099#112:2:A> And **they**_i are regularly published in the newspaper
 <ICE-IND:S1A-099#113:2:C> And others think you write \emptyset_i
 - (11) <ICE-SIN:S1A-092#243:1:B> Do you know who **she**_i is
 <ICE-SIN:S1A-092#244:1:A> No \emptyset_i doesn't sound familiar

After measuring the saliency and tightness of the link between anaphoric element and antecedent, three levels of accessibility were distinguished: high, intermediate and low. Highly accessible referents were those that were salient and, if applicable,¹⁶ showed a tight link between the anaphor and the antecedent. Referents of intermediate accessibility were those that were either salient or had a tight link. Finally, those referents that did not satisfy any of the conditions were considered to have low accessibility.

16. Some instances were only analyzed with respect to the saliency of their referents. These were pronouns referring to participants in the conversation (or the writer or reader in written texts). They are considered to have exophoric reference, that is, they refer directly to an entity in the situational context, so they do not have any linguistic antecedent.

6. Results

After the retrieval process, 693 omitted and 9,286 overt pronouns were identified. The distribution of the instances per variety, syntactic function, and referentiality is set out in Table 2. With a total percentage of 8.7 in SgE, 6.8 in IndE, and 5.3 in BrE, pronoun omission is not particularly widespread in any of the three varieties, but there are some differences between them: pronoun omission is more frequent in SgE than in BrE, with IndE occupying an intermediate position, and these results are statistically significant ($\chi^2(2) = 30.19, p < .001$). These differences remain constant irrespective of the function and referentiality of the pronoun, with the exception that IndE occupies the first position with direct object referential pronouns. In general, omitted pronouns are more frequent when they are referential, and when they are in subject rather than in direct object position (with the aforementioned exception in IndE).

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

Function		Variety					
		SgE		IndE		BrE	
		Omitted	Overt	Omitted	Overt	Omitted	Overt
Subject	Referential	9.4% (268)	90.6% (2,572)	6.5% (166)	93.5% (2,378)	6.4% (180)	93.6% (2,647)
	Non-referential	2.7% (7)	97.3% (253)	0.9% (2)	99.1% (220)	0% (0)	100% (264)
Direct Object	Referential	7.2% (24)	92.8% (308)	12.8% (44)	87.2% (300)	0.6% (2)	99.4% (315)
	Non-referential	0% (0)	100% (10)	0% (0)	100% (8)	0% (0)	100% (11)
Total		8.7% (299)	91.3% (3,143)	6.8% (212)	93.2% (2,906)	5.3% (182)	94.7% (3,237)

As for the distribution of omitted and overt pronouns in the three levels of accessibility,¹⁷ Table 3 shows that the percentage of omitted pronouns is higher in all the varieties when their referents are highly accessible. Most instances occur at this level of accessibility, with only a few cases at the intermediate and low levels. The distribution of overt pronouns is not radically different from that of omitted ones, with most instances occurring with highly accessible referents. This is not

17. Non-referential (overt and omitted) pronouns do not have referents and only fulfill a structural function in the clause, so they cannot be taken into account in an analysis dealing with the accessibility of entities in the discourse. Therefore, they were excluded from further analysis.

surprising because, according to Ariel's (2001) accessibility marking scale, both types of pronouns are markers of high accessibility.¹⁸ However, there are statistically significant differences in the distribution of omitted and overt pronouns with respect to the accessibility of their referents in the three varieties (SgE: $\chi^2(2) = 11.35$, $p < .01$; IndE: $\chi^2(2) = 7.1$, $p < .05$; BrE: $\chi^2(2) = 16.15$, $p < .001$). The level of accessibility thus influences the frequency of omitted pronouns to a higher degree than that of overt ones. As for differences between the varieties, SgE has the highest percentage of omission with highly accessible referents, IndE holds an intermediate position, and BrE has the lowest percentage of omission with this type of referents ($\chi^2(2) = 21.76$, $p < .001$). If we consider each accessibility factor individually, saliency plays a role in the distribution of omitted and overt pronouns in two out of three varieties: omitted pronouns are more frequent with salient referents in the three varieties, as can be seen in Table 4, but the differences are only statistically significant in SgE and in BrE (SgE: $\chi^2(1) = 8.51$, $p < .01$; IndE: $\chi^2(1) = 2.66$, $p > .05$; BrE: $\chi^2(1) = 8.17$, $p < .01$). The majority of the instances of both types of pronouns occur with salient referents, again due to the fact that they are both high accessibility markers, but saliency seems to influence the occurrence of omitted pronouns more strongly. There are statistically significant differences between the varieties, since the percentage of omitted pronouns with salient referents is higher in SgE, then in IndE, and finally in BrE ($\chi^2(2) = 24.68$, $p < .001$).

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

Accessibility	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
High	9.8%	90.2%	7.8%	92.2%	6.4%	93.6%
	(270)	(2,474)	(191)	(2,251)	(175)	(2,548)
Intermediate	6.4%	93.6%	4.3%	95.7%	2.4%	97.6%
	(18)	(263)	(12)	(267)	(7)	(282)
Low	2.7%	97.3%	4.2%	95.8%	0%	100%
	(4)	(143)	(7)	(160)	(0)	(132)

Focusing on the tightness of the link between the anaphoric element and its antecedent, we encounter more noticeable differences between omitted and overt pronouns. Table 5 shows that omission increases from 4.3 percent when the link is loose to 13.1 percent when it is tight in SgE ($\chi^2(1) = 17.28$, $p < .001$), from 3.6 percent to 13.4 percent in IndE ($\chi^2(1) = 22.4$, $p < .001$), and from 0 percent to 8.8 percent in BrE ($\chi^2(1) = 22.53$, $p < .001$). There are again differences between the

18. See Section 2.

Table 4. Distribution of overt and omitted pronouns with respect to the saliency of their antecedents

Saliency	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
Salient	9.7% (278)	90.3% (2,590)	7.5% (194)	92.5% (2,376)	6.2% (175)	93.8% (2,654)
Non-salient	4.6% (14)	95.4% (290)	5% (16)	95% (302)	2.2% (7)	97.8% (308)

varieties: SgE and IndE show higher percentages of omission than BrE both when the link between the anaphor and its antecedent is tight ($\chi^2(2) = 12.3, p < .01$) and when it is loose ($\chi^2(2) = 9.32, p < .01$). The factors that influence the tightness of the link affect omitted pronouns to different degrees, and omission seems to be favored in one syntactic context particularly. As shown in Table 6, omitted pronouns are very frequent when the anaphoric element and the antecedent occur in coordinated clauses (see Example 6 above), in all three varieties (SgE: $\chi^2(2) = 266.18, p < .001$; IndE: $\chi^2(2) = 224.66, p < .001$; BrE: $\chi^2(2) = 307.46, p < .001$). The percentage of omission is much lower when anaphor and antecedent occur in independent clauses, and almost negligible when they are in embedded clauses. Finally, pragmatic cohesion and the distance in clauses between the anaphoric element and its antecedent also play an important role in the distribution of omitted and overt pronouns (see Tables 7 and 8, respectively). In all three varieties, omission increases in pragmatically cohesive contexts¹⁹ and when there is a short distance (between 1 and 4 clauses) between the anaphoric element and its antecedent.²⁰

Table 5. Distribution of overt and omitted pronouns with respect to the tightness of the link between the anaphor and its antecedent

Link	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
Tight	13.1% (129)	86.9% (854)	13.4% (135)	86.6% (874)	8.8% (82)	91.2% (853)
Loose	4.3% (12)	95.7% (269)	3.6% (11)	96.4% (292)	0% (0)	100% (239)

19. SgE: $\chi^2(1) = 60.35, p < .001$; IndE: $\chi^2(1) = 70.17, p < .001$; BrE: $\chi^2(1) = 50.77, p < .001$.

20. SgE: $\chi^2(2) = 7.7, p < .05$; IndE: $\chi^2(2) = 9.98, p < .01$; BrE: $\chi^2(2) = 12.23, p < .01$.

Table 6. Distribution of overt and omitted pronouns with respect to the syntactic relation between the anaphor and its antecedent

Syntactic relation	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
Embedding	1.6% (3)	98.4% (186)	2.9% (5)	97.1% (169)	1% (2)	99% (208)
Coordination	50% (81)	50% (81)	46.6% (82)	53.4% (94)	41.1% (67)	58.9% (96)
Independence	6.5% (52)	93.5% (744)	7.1% (57)	92.9% (751)	1.8% (13)	98.2% (694)

Table 7. Distribution of overt and omitted pronouns with respect to the pragmatic cohesion between the anaphor and its antecedent

Pragmatic cohesion	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
Cohesive	22.8% (83)	77.2% (281)	23.7% (90)	76.3% (289)	15.4% (58)	84.6% (319)
Non-cohesive	6.9% (55)	93.1% (742)	6.7% (54)	93.3% (752)	3.4% (24)	96.6% (685)

Table 8. Distribution of overt and omitted pronouns with respect to the distance between the anaphor and its antecedent

Distance	Variety					
	SgE		IndE		BrE	
	Omitted	Overt	Omitted	Overt	Omitted	Overt
1–4 clauses	12.9% (125)	87.1% (846)	13.4% (135)	86.6% (874)	8.7% (81)	91.3% (852)
5–8 clauses	10.2% (10)	89.8% (88)	6.6% (6)	93.4% (85)	1.3% (1)	98.7% (74)
More than 8 clauses	3.3% (3)	96.7% (89)	3.5% (3)	96.5% (82)	0% (0)	100% (78)

To sum up, omitted pronouns occur more frequently when they refer to highly accessible antecedents than to those with intermediate or low levels of accessibility, and this was true for SgE, IndE, and BrE. Each of the accessibility factors, saliency and the tightness of the link between the anaphor and its antecedent, influence the

distribution of omitted and overt pronouns, so that omission is more common with salient referents and when the link is tight. Moreover, coordination is found to be a particularly favorable context for the occurrence of omitted pronouns. The results also show differences between the varieties, the most important one being that pronoun omission is more frequent in SgE, then in IndE, and finally in BrE. The next section will address these findings and relate them to the theoretical background of the paper and the research questions.

7. Discussion

Pronoun omission is a restricted phenomenon in all three varieties of English under analysis, with 8.7 percent in SgE, 6.8 percent in IndE, and 5.3 percent in BrE. The great majority of instances are overt, which means that, even though pronoun omission increases the efficiency of linguistic structures,²¹ the preferred option is to refer to entities by means of a (minimal, i.e. short) explicit marker. This suggests that, on most occasions, speakers consider that an overt form is necessary to guide the addressee to the correct antecedent. Omitted pronouns are basically restricted to contexts in which their referents can be recovered with ease: of the 684 instances of omitted pronouns with referents found in the texts, only 48 did not occur with highly accessible referents. There is a clear pattern of distribution: omission increases and reaches its peak with highly accessible referents and decreases at intermediate and, especially, low levels of accessibility. Therefore, the accessibility of entities in discourse influences the frequency of pronoun omission in SgE, IndE, and BrE alike. With respect to each of the individual factors, the distribution of omitted and overt pronouns is affected by the saliency of the referent and by the tightness of the link between the anaphoric element and its antecedent, so that omitted pronouns increase in frequency when they refer to salient referents and the link between the elements is tight. Of all the conditions that strengthen the tightness of the link, coordination influences pronoun omission to the greatest extent: when occurring in coordinate clauses, pronouns are omitted almost half the time. The other two conditions, pragmatic cohesion and a short distance in clauses between the anaphor and its antecedent, also increase the frequency of omitted pronouns, although to a lesser degree than coordination does.

If we compare the results obtained for SgE, IndE, and BrE, we encounter some important differences. The clearest one is that the total percentage of omission is not the same in the three varieties: it is higher in the high-contact varieties (that

21. See Section 3.2.

is, SgE and IndE) than in BrE, which supports claims that pronoun omission is a simplifying phenomenon and that contact simplifies languages (and dialects). As mentioned in Section 4, SgE and IndE are spoken in communities where non-native acquisition and use of English are the norm. BrE, however, is a native variety of English, and therefore it is not affected by language contact to the same extent as SgE and IndE. The results of the present study then reflect the simplifying effect of contact, since pronoun omission is more frequent in the high-contact varieties than in BrE. This increase in frequency in the high-contact varieties is not random – it is mostly restricted to pronouns that refer to highly accessible antecedents – and this is where the concept of communicative efficiency becomes important: formal minimization may simplify linguistic structures (and thus be easier for the speaker to articulate), but this is useless if the message that the structures convey cannot be decoded by the addressee. What happens in SgE and in IndE is that speakers omit pronouns more frequently than BrE speakers, but they do so almost exclusively when pronouns refer to highly accessible antecedents, thus keeping communication efficient.²² The differences between SgE and IndE, on the one hand, and BrE, on the other, are not only quantitative; they are also qualitative due to the referentiality and the grammatical function of the pronouns omitted. Omission with non-referential pronouns does not occur in BrE, and it is almost negligible in direct object position. Omitted pronouns are thus not only more frequent in SgE and in IndE than in BrE, they also occur in contexts in which they are not commonly found in the native variety.

IndE and SgE have different distribution patterns of omitted and overt pronouns. Pronoun omission is more frequent in SgE than in IndE, and accessibility seems to have a stronger influence in the former than in the latter variety. As mentioned in Section 5, even though SgE and IndE are high-contact varieties of English, they belong to different subtypes: IndE is an indigenized L2 and SgE a shift L1 variety. Due to these statuses, there is a difference in the attitudes towards the variety of English.²³ In India, English is used mostly for specific purposes and domains of life, including government and administration, politics, higher education, etc. It is mainly a second language, and the grammatical innovations, characteristic of IndE, are still considered by many to be mistakes. In Singapore, on the other hand, English is becoming a first language and increasingly used by many people in more

22. The reader may have noticed that Table 3 also shows higher percentages of omission in SgE and IndE than in BrE at the intermediate and low levels of accessibility. These differences, however, are not statistically significant (intermediate: $\chi^2(2) = 5.43$, $p > .05$; low: $\chi^2(2) = 5.93$, $p > .05$).

23. For a detailed description of the evolution and present state of the attitudes of IndE and SgE speakers towards their varieties, see Schneider (2007: 153–173).

intimate domains of life. SgE, with all its particularities, carries positive connotations as a marker of Singapore's national and cultural identity, and its grammatical innovations are accepted as part of the correct norm. The negative attitude towards IndE and the positive one towards SgE may have inhibitory and facilitating effects, respectively, on the features that differentiate them from the grammars of other varieties, which could account for the higher frequency of pronoun omission in SgE than in IndE. However, this is a rather ad-hoc explanation; further research is needed to clarify this issue.

The results presented in the previous section evidence that the accessibility of referents is a variable that affects pronoun omission in SgE and IndE (and also in BrE), that is, omitted pronouns refer almost exclusively to highly accessible entities or events. However, with a frequency of omission with highly accessible referents of 9.8 percent in SgE and 7.8 percent in IndE (versus 90.2 percent and 92.2 percent, respectively, for overt pronouns), accessibility preferences cannot be said to have been conventionalized in either variety: it is not a rule of the grammars of SgE and IndE that highly accessible referents are marked only (or even often) by omitted pronouns. They are generally coded by overt pronouns, although on a few occasions these pronouns are omitted. There is one exception here: coordinate clauses. In this context, pronouns are omitted in half the cases in SgE and in almost half (46.6 percent) in IndE, which makes omitted pronouns a common choice and, possibly, a conventionalized one. Many accessibility factors apply in coordinate clauses at the same time: in most cases the antecedent is in the immediately preceding clause (except when some of the conjuncts contain embedded clauses within them), in most cases the antecedent is also in subject position (although there are some cases in which it is in other positions),²⁴ and coordinate clauses are connected by a conjunction, which increases their pragmatic cohesion. All of these facilitating factors, together with the fact that pronouns are also very frequently omitted in coordination in BrE (in 41.1 percent of the cases), make coordinate clauses a perfect context for omission.

Hawkins' Performance Grammar Correspondence Hypothesis suggests that users prefer those grammatical features that increase the efficiency of structures, and they end up becoming conventionalized in the grammars of languages. It was shown that pronoun omission increases the efficiency of structures when the referent of the omitted pronoun is accessible and that second language users of English (and other languages) drop pronouns frequently. It was then attested that, in high-contact varieties of English, that is, those with high numbers of non-native speakers in the past or present, pronoun omission is more frequently found than in

24. See Example 3.

low-contact varieties. Finally, two high-contact varieties of English, IndE and SgE, omitted pronouns more frequently than BrE and mostly refer to highly accessible referents, though the most common choice was an overt pronoun. In coordinate clauses, however, pronoun omission is a common option. These findings suggest that high accessibility in general is not conventionalized in the grammars of either IndE or SgE as a context in which pronoun omission is the required or even the preferred option, but that coordinate clauses may occur in such a context, which provides further support for Hawkins' hypothesis.

8. Conclusions

The present article set out to achieve three different goals. The first was to study the complexity and communicative efficiency of pronoun omission. It was argued that such omission results in simpler linguistic structures but that it also increases their efficiency if referents of omitted pronouns are accessible. The second goal was to examine whether language contact influences the occurrence of pronoun omission in varieties of English. It was found that this grammatical feature is more common in high-contact varieties than in low-contact ones, which is further evidence that contact fosters simplification. Moreover, the results of a corpus study revealed that pronoun omission is more frequent in SgE and IndE, two high-contact varieties, than in BrE, a low-contact one. Finally, the third goal here was to analyze the distribution of omitted and overt pronouns with respect to the accessibility of their referents in SgE, IndE, and BrE. I also explored whether pronoun omission is a conventionalized feature of the grammars of the high-contact varieties, in accordance with Hawkins' Performance Grammar Correspondence Hypothesis. It was found that omitted pronouns almost exclusively referred back to highly accessible antecedents and that coordination was a particularly favorable context for omission in both varieties. However, it was argued that overt pronouns were still the default option for antecedents of this type, which means that omitted pronouns are not a conventionalized marking strategy to code high accessibility either in IndE or SgE. As for coordination, on the other hand, omission was a very common alternative in both varieties, and possibly a conventionalized one, which provides support for Hawkins' hypothesis. In addition, some differences between SgE and IndE were found: pronoun omission was more frequent in the former than in the latter. These distinct distribution patterns were explained by the fact that the varieties belong to different types: IndE is an indigenized L2 variety and SgE is becoming a shift L1 variety.

To conclude, the present article has provided interesting findings on the distribution of omitted pronouns in varieties of English in general and, more specifically, in IndE and SgE. However, some potential shortcomings must be recognized. Ideally, a bigger corpus could have been used, but the retrieval of pronoun omission had to be done manually, which was very time-consuming. For practical reasons, then, the analysis was restricted to the set of texts mentioned in Section 6, which provided a corpus of approximately 130,000 words. Despite its moderate size, this corpus yielded a fairly large number of instances of omitted and overt pronouns, and thus it can be assumed that the results of the present study are representative. Another issue to be borne in mind is that the high-contact varieties of English that have been examined here, SgE and IndE, have substrate languages in which pronouns can be omitted. It would also be interesting to analyze other varieties with non-pro-drop substrate languages in order to disentangle simplifying processes from transfer effects. This issue, however, must be left for future research.

Acknowledgements

For generous financial support, I am grateful to the European Regional Development Fund and the following institutions: Regional Government of Galicia (Directorate General for Scientific and Technological Promotion, grant GPC 2014/004); Spanish Ministry of Economy and Competitiveness (grants FFI2014-51873-REDT, FFI2014-52188-P and BES-2015-071 233). Thanks are also due to Teresa Fanego and J. Carlos Acuña-Fariña, and to Daniel Schreier and two anonymous reviewers of *English World-Wide* for their most valuable suggestions on earlier versions of this paper.

Sources

- 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, and Kerstin Lunkenheimer, 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).

References

- Ariel, Mira. 1988. "Referring and Accessibility". *Journal of Linguistics* 24: 65–87.
doi:10.1017/S0022226700011567
- 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: 3–42. doi:10.1017/S0022226700016170
- Ariel, Mira. 2001. "Accessibility Theory: An Overview". In Ted Sanders, Joost Schliperoord, and Wilbert Spooren, eds. *Text Representation*. Amsterdam: Benjamins, 29–87.
doi:10.1075/hcp.8.04ari
- Bao, Zhiming. 2001. "The Origins of Empty Categories in Singapore English". *Journal of Pidgin and Creole Languages* 16: 275–319. doi:10.1075/jpcl.16.2.03zhi
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Cole, Melvyn D. 2009. "Null Subjects: A Reanalysis of the Data". *Linguistics* 47: 559–587.
doi:10.1515/LING.2009.019
- Cole, Melvyn D. 2010. "Thematic Null Subjects and Accessibility". *Studia Linguistica* 64: 271–320.
doi:10.1111/j.1467-9582.2010.01172.x
- Dahl, Östen. 2009. "Testing the Assumption of Complexity Invariance: The Case of Elfdalian and Swedish". In Geoffrey Sampson, David Gil, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 50–63.
- Fenk-Oczlon, Gertraud, and August Fenk. 2008. "Complexity Trade-offs between Subsystems of Language". In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, eds. *Language Complexity: Typology, Contact, Change*. Amsterdam: Benjamins, 43–65. doi:10.1075/slcs.94.05fen
- Gundel, Jeanette, and Elaine Tarone. 1983. "Language Transfer and the Acquisition of Pronominal Anaphora". In Susan Gass and Larry Selinker, eds. *Language Transfer and Language Learning*. Rowley: Newbury House, 281–296.
- Gundel, Jeanette, Nancy Stenson, and Elaine Tarone. 1984. "Acquiring Pronouns in a Second Language: Evidence for Hypothesis Testing". *Studies in Second Language Acquisition* 6: 215–225. doi:10.1017/S0272263100005027
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
doi:10.1093/acprof:oso/9780199252695.001.0001
- Hawkins, John A. 2009. "An Efficiency Theory of Complexity and Related Phenomena". In Geoffrey Sampson, David Gil, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 252–268.
- Heine, Bernd, and Tania Kuteva. 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511614132
- Huang, C.-T. James. 1984. "On the Distribution and Reference of Empty Pronouns". *Linguistic Inquiry* 15: 531–574.
- Huang, Yan. 1992. "Against Chomky's Typology of Empty Categories". *Journal of Pragmatics* 17: 1–29. doi:10.1016/0378-2166(92)90026-8
- Huang, Yan. 2000. *Anaphora: A Cross-linguistic Study*. Oxford: Oxford University Press.
- Huddleston, Rodney, Geoffrey K. Pullum, Laurie Bauer, Betty Birner, Ted Briscoe, Peter Collins, David Denison, David Lee, Anita Mittwoch, Geoffrey Nunberg, Frank Palmer, John Payne, Peter Peterson, Lesley Stirling, and Gregory Ward. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Jaeggli, Osvaldo, and Kenneth J. Safir, eds. 1989. *The Null Subject Parameter*. Dordrecht: Kluwer.
doi:10.1007/978-94-009-2540-3

- Kusters, Wouter. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.
- McWhorter, John H. 2007. *Language Interrupted: Signs of Non-native Acquisition in Standard Language Grammars*. New York: Oxford University Press.
doi:10.1093/acprof:oso/9780195309805.001.0001
- Miestamo, Matti. 2008. "Grammatical Complexity in a Cross-linguistic Perspective". In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, eds. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 23–41. doi:10.1075/slcs.94.04mie
- Miestamo, Matti, Kaius Sinnemäki, and Fred Karlsson, eds. 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins. doi:10.1075/slcs.94
- Moag, Rodney F., and Robert Poletto. 1991. "Discourse Level Evidence for South Asia as a Linguistic Area". *Language Sciences* 3: 229–254. doi:10.1016/0388-0001(91)90016-T
- Neeleman, Add and Kriszta Szendrői. 2007. "Radical Pro Drop and the Morphology of Pronouns". *Linguistic Inquiry* 38: 671–716. doi:10.1162/ling.2007.38.4.671
- Nichols, Johanna. 2009. "Linguistic Complexity: A Comprehensive Definition and Survey". In Geoffrey Sampson, David Gil, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 110–125.
- Rizzi, Luigi. 1986. "Null Objects in Italian and the Theory of *pro*". *Linguistic Inquiry* 17: 501–558.
- Sampson, Geoffrey. 2009. "A Linguistic Axiom Challenged". In Geoffrey Sampson, David Gil, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 243–251.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618901
- Scott, Mike. 2012. *WordSmith Tools*. Version 6. Liverpool: Lexical Analysis Software.
- Sinnemäki, Kaius. 2008. "Complexity Trade-offs in Core Argument Marking". In Matti Miestamo, Kaius Sinnemäki, and Fred Karlsson, eds. *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 67–88. doi:10.1075/slcs.94.06sin
- Szmrecsanyi, Benedikt, and Bernd Kortmann. 2009. "Between Simplification and Complexification: Non-standard Varieties of English around the World". In Geoffrey Sampson, David Gil, and Peter Trudgill, eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 65–79.
- Trudgill, Peter. 2009a. "Sociolinguistic Typology and Complexification". In Geoffrey Sampson, David Gil, and Peter Trudgill, 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 Markku Filppula, Juhani Klemola, and Heli Paulasto, eds. *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. London: Routledge, 304–322.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.
- Williams, Jessica. 1987. "Non-native Varieties of English: A Special Case of Language Acquisition". *English World-Wide* 8: 161–199. doi:10.1075/eww.8.2.02wil
- Williams, Jessica. 1988. "Zero Anaphora in Second Language Acquisition: A Comparison among Three Varieties of English". *Studies in Second Languages Acquisition* 10: 339–370.
doi:10.1017/S0272263100007488
- Zyzyk, Eve. 2008. "Null Objects in Second Language Acquisition: Grammatical vs. Performance Models". *Second Language Research* 24: 65–110. doi:10.1177/0267658307082982

Author's address

Iván Tamaredo
University of Santiago de Compostela
Department of English and German
Avenida de Castelao (Campus Norte)
15782 Santiago de Compostela
Spain
ivan.tamaredo@usc.es