Pronoun omission and agreement: An analysis based on ICE Singapore and ICE India

Iván Tamaredo and Teresa Fanego, University of Santiago de Compostela

Abstract

This article deals with pronoun omission in subject position and its connection with subject-verb agreement in Indian English and Singapore English. Agreement morphology has been found to be a predictor and facilitator of pronoun omission cross-linguistically in that it aids in the identification and retrieval of the referents of omitted pronouns. The results of a corpus study partly confirm this trend, since they show that agreement morphology does have a weak facilitating effect in both varieties examined; that is, pronoun omission increases when the subject and the verb agree in person and number. However, this is only true for lexical verbs; non-modal auxiliaries (i.e., be, have, do*), on the contrary, show a low percentage of omitted pronouns and no facilitating effect of agreement morphology. To account for this finding, the possible inhibiting effect on pronoun omission of the frequency of co-occurrence of pronouns and non-modal auxiliaries was also explored.*

1   Introduction

Pronoun omission in subject position (also called pro-drop, null subject parameter, or zero anaphora) has often been correlated cross-linguistically with agreement morphology (cf. Corbett 2006), because it tends to occur in languages with rich subject-verb agreement. Poor agreement morphology, on the other hand, is too weak to guide the addressee in the identification of the referents of omitted pronouns, and thus is not generally considered to be a licensor of pronoun omission (cf. Speas 1994, 2006). Against this view, however, it has also been pointed out that there are languages that lack agreement entirely and yet allow subjects to be omitted (cf. Huang 1984; Jaeggli and Safir 1989). Likewise, as will be shown in the present article, some languages with poor agreement morphology, such as the varieties of English spoken in India and Singapore (Bhatt 2004; Wee 2004), can also exhibit pronoun omission. The picture that emerges from these

varied observations is that agreement is a sufficient but not a necessary condition for the occurrence of null subjects.

In the present paper, it will be argued that the relation between agreement morphology and pronoun omission can be better explained by an inclusive approach that views agreement as a formal cue that helps recover the referents of omitted pronouns, but also acknowledges the role of co-text or context in the recoverability of omitted referents. Such an approach has been formulated by Cole (2009, 2010), and its application here to Indian English and Singapore English serves to explain, largely, the pattern of distribution of omitted pronouns with respect to subject-verb agreement in those varieties. This said, some cases of pronoun omission in Indian English and Singapore English remain which Cole's theory does not seem to cover, so a hypothesis will be put forward to explain those cases on the basis of the effect of frequency of co-occurrence on the processing and storage of linguistic units.

The paper is structured as follows. Section 2 reviews the literature on pronoun omission and its connection with agreement. Section 3 deals with the status of the varieties examined in their respective speech communities and discusses the linguistic substrata of India and Singapore. Section 4 explains the data and methodology used in the study. In Section 5 the results of the data analysis are presented, followed by a discussion of these in Section 6. Section 7 concludes the paper.

2   *Pronoun omission and agreement*

In the present article, the term 'pronoun omission' is used to refer to those cases in which there is a gap in the subject position of a clause that could be filled by an overt personal pronoun. Both referential and non-referential or dummy pronouns (*it* and *there*) are included in this definition, as can be seen in examples (1) and (2) respectively, taken from the Indian (ICE-IND) and Singaporean (ICE-SIN) components of the *International Corpus of English* (ICE):

(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]$_i$** <,>
<ICE-IND:S1B-055#12:1:C> The honourable minister has made a statement saying within three years $\emptyset_i$ have increased to <,> up to sixty percent of the last <,> pre previous production <,>

(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 Ø will be [...] on made easier for her to move around

Instances of omission in non-finite clauses were not taken into account because they belong to a different class of empty elements (see below). Omitted subjects in imperative clauses were also excluded from the study because their occurrence is independent from subject-verb agreement.

2.1   Literature review

The phenomenon of pronoun omission has been widely studied and analysed cross-linguistically, either with a focus mainly on its formal and structural characteristics (Chomsky 1981, 1982; C.-T. J. Huang 1984; Rizzi 1986; Jaeggli and Safir 1989; Neeleman and Szendrői 2007) or from a pragmatic and/or cognitive perspective (Y. Huang 1992, 2000; Cole 2009, 2010; cf. also Ariel 1988, 1990, 1994, 2001 for a cognitive theory of anaphoric expressions). The remainder of this section will present a brief review of the literature, paying special attention to the connection between pronoun omission and agreement.

Most approaches within the generative tradition recognize the existence of empty elements that fulfil a function from the point of view of syntax but do not have an overt phonetic realization. Chomsky (1981, 1982) put forward a theory which accounted for these elements within the Government and Binding framework. This account evolved following new developments in generative linguistics, but the basic ideas remained unchanged. Four different classes of empty elements are distinguished: NP-traces, pro, PRO, and *wh*-traces or variables. PRO and pro are generated *in situ* in the position they occupy: pro as a subject in finite clauses and PRO in the subject position of non-finite clauses. NP-traces and variables are left behind by movement operations such as raising or topicalization and are coindexed with the moved element. The difference between them is that NP-traces are coindexed with an element that moves to an A-position (i.e., subject or object), the antecedent of a variable moves to an A'- or non-thematic position such as a sentence topic (cf. Haegeman 1994 for a summary of the different types of empty elements and their distribution within Government and Binding theory). In what follows we will concentrate on the elements pro

and variable and their connection with verbal agreement, which is the main focus of the present article.

Most frameworks within generative grammar postulate that there must always be an element in the subject position of a clause, either an overt or a non-overt one. The empty element that fulfils this function in the case of finite clauses is pro, which has definite reference with its antecedent being located outside the clause, in the co-text or in the context, as in the case of overt pro-nouns. Null subjects are not allowed in all languages; in some of them the subject position of a finite clause has to be always filled by an overt element. One of the factors that has been identified, from the early days of pro-drop theory, as licensing the occurrence of empty elements, is agreement between the verb and the null constituent: languages with rich agreement between the verb and one or more of its arguments often allow empty constituents (Corbett 2006; cf. also Taraldsen 1980). Rizzi (1986) developed the following account of the restrictions on the occurrence of pro and its connection with agreement: pro is licensed by a governing head, which in rich agreement languages is the verbal inflection, and its semantic content is retrieved from the formal features of this head. This explains why many languages with poor or no agreement do not allow the occurrence of pro, because its referent cannot be identified from the underspecified or nonexistent agreement affixes on the verb. However, this is not so in all cases and, as pointed out by C.-T. J. Huang (1984), there are languages with no agreement that allow the occurrence of empty elements in subject and even in object position. In Chinese, for instance, empty elements are allowed when they refer either to an element in the same sentence that functions as a governing head or to a discourse topic. The former are cases of pro (3) while the latter are variables coindexed with a null sentence topic (4). In (3) (example 36 in Huang 1984: 542), the empty subject of the subordinate clause is a pro that is coindexed with the subject of the main clause, i.e., a person named Zhangsan. Zhangsan acts as the governing head that licenses the occurrence of pro in this sentence and provides the referent for the omitted element. On the other hand, in (4), which shows a second possible interpretation of the structure in (3), the empty element is a variable that refers to a previously mentioned antecedent: the variable occupies the subject position of the subordinate clause and is coindexed with a null sentence topic ($[_{\text{Topic}} \varnothing_i]$) that would in turn refer to an entity fixed in discourse; i.e., an individual different from Zhangsan and who has already been mentioned in the conversation.

(3) Zhangsan$_i$ shuo **Ø$_i$** mingtian yao lai.
     Zhangsan say tomorrow want come
     Zhangsan$_i$ said that he$_i$ wants to come tomorrow.

(4) [$_{Topic}$ Ø$_i$] Zhangsan shuo **Ø$_i$** mingtian yao lai.
     Zhangsan say tomorrow want come
     [$_{Topic}$ Ø$_i$] Zhangsan said that he/she$_i$ wants to come tomorrow.

C.-T. J. Huang's analysis of empty elements in Chinese has been called into question because, as Y. Huang (1992) argues, it does not provide a valid formal mechanism that explains in which cases pro is licensed and how it retrieves its content from the subject of the main clause. Additionally, the issue of how null topics are identified with their antecedents is not a syntactic but a pragmatic one. Y. Huang (1992: 27) concludes that empty elements in Chinese are "realized by syntactically undifferentiated gaps" which "are not grammatically but pragmatically determined." Other purely syntactic explanations of the distribution of omitted pronouns in languages with poor agreement, or with no such agreement, have been put forward. Neeleman and Szendrői (2007) argue that omission occurs in such languages only if they have pronouns with agglutinative morphology. However, there are several counterexamples to this theory: the varieties of English that are the topic of the present article, for example, have pronouns with fusional morphology but allow their omission.

2.2 Cole's (2009, 2010) theory of pronoun omission

From the brief review presented in Section 2.1, we can conclude that, on the one hand, agreement is an important licensor of pronoun omission, since languages with rich agreement often allow the occurrence of omitted pronouns. On the other hand, it is not a necessary condition, because empty elements can also be found in languages with poor or no agreement. It seems that, in this latter group of languages, pragmatic considerations may be more important for the licensing of omitted pronouns, since there appears to be no strictly syntactic mechanism that determines their occurrence. What we need, then, is a theory that can unite both factors in a single coherent explanation of pronoun omission. Cole (2009, 2010) has proposed the hypothesis that both agreement and the accessibility of referents in discourse affect the occurrence of omitted pronouns. He argues that the decision to omit a pronoun or to leave it overt is a process that adheres to a set of rules. Thus, a pronoun may be omitted (Cole 2009):

- If agreement affixes on the verb allow the identification of the morphological features of the omitted pronoun (i.e., person, number and gender) up to the point of a language's 'morphological maximality'.
- If agreement is not enough to recover all or some of the features, then there must be an accessible antecedent in context from which the omitted pronoun can recover them.
- If agreement is not enough and there is not an accessible antecedent in context, the omitted pronoun can be assigned a preferred interpretation in some languages, e.g., first person singular in Italian.
- If none of the above is true, then an overt pronoun must be used.

A language's morphological maximality is determined by the set of features and feature combinations that can be unambiguously recovered from the agreement affixes attached to the verb. In English, for example, these would be person and number, but only in the present tense and with certain person-number combinations: third person singular with most verbs (except modals), and first and third person singular with the verb *be* (*am* and *is*). In all other cases, either there is no subject-verb agreement, or the forms of the verb are not enough to identify the person and number of the subject.[1] Regarding the accessibility of the antecedent in the context, an entity or an event is highly accessible when it is salient in discourse, for example when it is a discourse topic, and whether there is a close connection between the antecedent and the anaphoric element (cf. Ariel 1988, 1990, 1994, 2001 for an account of the factors that affect the accessibility of an antecedent in the discourse).

In a language such as Chinese, with no agreement between the subject and the verb, morphology cannot help in the identification of the features of an omitted pronoun and, therefore, there must be an accessible antecedent in context. In English (i.e., in the varieties that are the focus of the present article), where subject-verb agreement is limited to some person-number combinations, morphology can only be helpful in certain cases, and in the rest an accessible antecedent is required. In Spanish, a rich agreement language, morphology can be used to recover the person and number of the omitted pronoun in most cases, with the exception of those in which ambiguity arises due to syncretism, i.e., when the same form of the verb can refer to more than one person-number combination: *comía* in Spanish (*I/he/she/it ate*) can be used with a first person singular and a third person singular subject.

In Cole's theory (2009, 2010), agreement functions as an aid to recover the morphological features of an omitted pronoun that is only available in poor and rich agreement languages; yet even in the latter this is not enough in some cases.

Ultimately, all omitted pronouns must be coindexed with an accessible antecedent in all languages (even in those with rich subject-verb agreement); otherwise their referent cannot be correctly identified. However, agreement may facilitate the omission of pronouns in those languages in which it is available, and, as has been shown in this section, it is a powerful predictor of pronoun omission cross-linguistically. In the present article, we focus on the connection between agreement and pronoun omission in Indian English and Singapore English, two Asian varieties of English which allow the occurrence of omitted pronouns despite their poor agreement morphology. The next section turns to the varieties in question and their sociolinguistic background.

3 Indian English and Singapore English

English is considered to be a prototypical case of a language that does not allow the omission of pronouns. In present-day standard (British and American) varieties of English, subjects can occasionally be omitted in casual style when they occur at the beginning of main clauses, both declarative and interrogative (Quirk *et al*. 1985: 896–898; Huddleston and Pullum *et al*. 2002: 1540–1541), and also in instructional writing (Ruppenhofer and Michaelis 2010). Additionally, a pronoun can be covert in coordinate clauses when it is the subject of the second conjunct and it is coreferential with that of the first one. In other varieties, however, such as in Indian English and Singapore English, among others, omission occurs more frequently and in more contexts than those mentioned (cf. Bhatt 2004; Wee 2004). The rest of this section will first deal with the status of English in Indian and in Singaporean societies, and, second, with the local languages spoken in each country, which could have influenced the evolution of the varieties.

3.1 The status of English in India and Singapore

Despite the fact that nowadays both Indian English and Singapore English are recognized as varieties of English in their own right, their statuses in their respective speech communities are very different. The use of English in India is mostly restricted to specific domains of life, such as the government, administration, politics, higher education, the legal system, business and the media (Schneider 2007). It also functions as an "interethnically neutral link language" (Schneider 2007: 167), but it is not a marker of identity. Indian English is a second language variety: out of the sixty million users of English (Bhatt 2004), only 226,449 are native speakers according to the 2001 census (Lange 2012). The current status of English in India has been described by Mukherjee (2007:

158) as a "steady state" in which both progressive and conservative forces are at play. The most significant progressive forces are the linguistic innovations that characterize Indian English and distinguish it from other varieties, but there are also certain attitudes and policies that point to the possibility of English becoming more established in India. Among these are the proliferation of literary works in English written by Indian authors, and the recent adoption of English as a compulsory subject in primary education. On the other hand, there are also conservative forces at play which prevent English from increasing its influence in the country and becoming a carrier of Indian identity. First of all, teachers of English follow predominantly a British norm, which works against the establishment of an Indian standard. Secondly, and most importantly, many Indian English speakers still have a critical and self-deprecating attitude towards the local variety and instead hold native varieties as the appropriate and correct ones.

The situation in Singapore is different. English is the language of politics, administration, the courts and education, but is also used in more informal and intimate contexts: in 2010, English was spoken at home by 32.3 percent of the population of Singapore (Leimgruber 2013). Singapore English is still predominantly a second language but it is increasingly being acquired natively by the younger generations. The attitude towards the local variety is very different from that of Indian speakers. Singapore English, with all its particular features, is not only accepted as a correct norm but functions as an identity carrier. After the 1970s, Singapore became an industrialized and economically prosperous country with a mixture of European and Asian influences (Schneider 2007); Singapore English in turn came to be inseparable from this new cultural and national identity, and as a result Singaporeans developed a positive attitude towards it.

3.2   Linguistic substrata

India is a highly multicultural and multilingual nation. The Constitution recognizes eighteen official languages (Mukherjee 2007), but many more are spoken in the country. The majority of these languages belong either to the Indo-Aryan or to the Dravidian families. Hindi, which is part of the Indo-Aryan family, has a special status alongside English as the language used for official purposes. It is the native language of approximately 35 percent of the Indian population, which makes it the language with the highest number of speakers in India, and therefore it is an important substrate language for Indian English. With respect to pronoun omission, it is commonly used as a device to establish cohesion in oral and written language in Hindi, but also in Dravidian languages such as Malayalam

and in other South Asian languages (Moag and Poletto 1991). Omitted pronouns tend to refer to known or established antecedents in discourse, and they are the unmarked option in conversation. Hindi also lacks non-referential pronouns such as English dummy *it* and *there* (Kachru 2006). As for subject-verb agreement, it is present in Hindi, but also in most of the other Indo-Aryan and Dravidian languages to some degree (Moag and Poletto 1991). Person and gender are marked in the verb in Hindi, and this formal cue is used to retrieve the referents of anaphoric elements, including those of omitted pronouns.

The majority of the population of Singapore descend from the Chinese, Malay and Indonesian immigrants that moved to the island in the nineteenth century after its annexation by the British in 1819 (Bao 2001). To begin with, people of Chinese and Malay descent were the major ethnic groups, and their proportions were very similar, but from the 1840s onwards the population of Chinese origin grew rapidly and became the largest. This distribution still applies today, and thus Chinese is the most important substrate language for Singapore English. The nineteenth-century Chinese immigrants that moved to Singapore spoke many different dialects, but since 1979 the government has advocated the use of Mandarin among the Chinese population (Schneider 2007). Today, it is one of the official languages of Singapore, together with Malay, Tamil and English. With respect to pronoun omission, entities that are understood from context may be omitted in Chinese (Li and Thomson 1989). Pronouns may also be non-overt when they refer to general or nonspecific antecedents. A common phenomenon in Chinese discourse is what Li and Thompson (1989: 659) call a "topic chain", i.e., when "a referent is referred to in the first clause, and then there follow several more clauses talking about the same referent but not overtly mentioning that referent". In general, overt pronouns are only used in those situations in which the speaker wants to emphasize or highlight an antecedent. Mandarin Chinese is an isolating language, which means that most of the morphological features found in fusional and agglutinative languages are not attested. Subject-verb agreement is one of these features that are not found in Chinese, although some of its functions are fulfilled by other linguistic resources, such as word order.

India and Singapore have different linguistic substrata, and this is bound to be reflected in the varieties of English spoken in each country. In both cases, pronoun omission is a feature of the grammars of the indigenous languages, and this may very well be the reason why it occurs at all in Indian English and Singapore English, even though its use is very restricted in the standard varieties. The situation changes, however, with subject-verb agreement: it is present in the grammars of many languages spoken in India but it is not attested in Chinese. If

the substrate languages have influenced the varieties in this respect, then the relation between pronoun omission and agreement may not be the same in Indian and Singapore English. Since Chinese does not have subject-verb agreement, it could be the case that in Singapore English it is not exploited as a means of resolving anaphoric relations as much as in Indian English, which has substrate languages in which agreement provides information about the formal features of omitted subjects. This issue is discussed in Section 6.

4   Data and methodology

The present study revolves around the following research question: is there an effect of agreement in Indian English and/or Singapore English such that, when the verb unambiguously signals the person and number of the subject, the percentage of omitted pronouns increases? In other words, does the presence of agreement facilitate the omission of pronouns in one or both varieties? Rich subject-verb agreement has been found to license pronoun omission cross-linguistically, but Indian English and Singapore English have poor agreement morphology which may not be strong enough to have a significant effect. If there were such an effect, however, this would mean that, even in the case of poor agreement languages, speakers use whatever resources are available, formal or otherwise, to help them recover the referents of omitted pronouns, in line with Cole's (2009, 2010) theory.

In order to answer the research question, verbs occurring with omitted and overt pronouns were classified into lexical verbs, non-modal auxiliaries, and modal auxiliaries (Huddleston and Pullum *et al.* 2002). Non-modal auxiliaries include the verbs *be*, *do* and *have*. Modal auxiliaries comprise *can*, *could*, *may*, *might*, *must*, *ought*, *shall*, *should*, *will* and *would*. If agreement had a widespread effect on pronoun omission, we would expect to find a higher percentage of omitted pronouns with non-modal auxiliaries than with lexical verbs, because more forms of the verb agree in person and number with the subject in the former than in the latter.[2] Additionally, we would expect a higher frequency of omission with lexical verbs than with modal auxiliaries, for similar reasons. Within each class of verbs, those forms displaying agreement (i.e., *am* and *is* for the verb *be*, and the third person singular present forms of the rest of the verbs) should also exhibit a higher percentage of omission than the forms with no agreement. Therefore, each instance of a verb was also classified as to whether it showed agreement with the subject or not. The data was then analysed using the chi-squared test of independence in order to look for significant correlations in

the distribution of omitted and overt pronouns. The statistical analysis of the data was performed using the software R (R Development Core Team 2015).

The data for the study was retrieved from a set of texts taken from ICE-IND and ICE-SIN. Forty texts were selected (approximately ninety thousand words); twenty per variety. The texts were carefully chosen taking into account register (informal versus formal) and mode (spoken versus written), so that the final sample was balanced. Thus, for each variety five texts were spoken and informal, five were spoken and formal, five were written and informal, and five were 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 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, C, D, and E categories. Instances of verbs in combination with overt pronouns were recovered from the corpus using WordSmith Tools (Scott 2012), but those occurring with omitted pronouns had to be retrieved manually.

5 Results

After the retrieval process, 166 omitted pronouns and 2,600 overt ones were identified in the set of texts taken from ICE-IND. In the case of ICE-SIN, 275 omitted and 2,819 overt pronouns were recovered from the texts. The distribution of omitted and overt pronouns per variety is set out in Table 1:

Table 1: Distribution of omitted and overt pronouns per variety

Variety	Pronoun	
	Omitted	**Overt**
Indian English	6% (166)	94% (2,600)
Singapore English	8.9% (275)	91.1% (2,819)

Table 1 shows that omission is more frequent in Singapore English (8.9 percent) than in Indian English (6 percent), and that the difference is highly significant (χ^2 (1) = 17.49, $p < 0.001$). With respect to the referentiality of the pronouns, omission is more frequent with referential pronouns than with non-referential ones in both varieties, as reflected in Table 2:

Table 2: Frequency of omission with referential and non-referential (*it, there*) pronouns

Referentiality	Variety			
	Indian English		**Singapore English**	
	Omitted	**Overt**	**Omitted**	**Overt**
Referential	6.4% (162)	93.6% (2,372)	9.4% (267)	90.6% (2,567)
Non-referential	1.7% (4)	98.3% (228)	3.1% (8)	96.9% (252)

In Indian English, there are fewer instances of non-referential omitted pronouns than referential ones, and this distribution is statistically significant (χ^2 (1) = 8.21, $p < 0.01$). The situation is similar in Singapore English, i.e., fewer cases of omission with non-referential pronouns (χ^2 (1) = 11.84, $p < 0.001$), but the omission percentages are higher than in Indian English, in line with the pattern outlined in Table 1.

Let us now turn to the distribution of omitted and overt pronouns in the three classes of verbs distinguished in Section 4, shown in Table 3:

Table 3: Distribution of omitted and overt pronouns per verb class

Verb Class	Variety			
	Indian English		**Singapore English**	
	Omitted	**Overt**	**Omitted**	**Overt**
Lexical Verbs	8.1% (89)	91.9% (1,009)	11% (145)	89% (1,170)
Non-modal Aux.	3.8% (47)	96.2% (1,202)	5.3% (68)	94.7% (1,227)
Modal Aux.	7.2% (30)	92.8% (389)	12.8% (62)	87.2% (422)

In Indian English, omitted pronouns are more frequent when they occur with lexical verbs (8.1 percent) than with non-modal auxiliaries (3.8 percent). The percentage of omission with modal verbs (7.2 percent) is also higher than with non-modal auxiliaries, but lower than with lexical verbs. This distribution is highly significant (χ^2 (2) = 20.71, $p < 0.001$). In Singapore English, there is a similar situation, but with two minor differences. First, omission is again more

frequent overall than in Indian English, in accordance with Table 1. Second, in Singapore English non-modal auxiliaries also have the lowest percentage of omitted pronouns (5.3 percent), but this time it is higher with modal auxiliaries (12.8 percent) than with lexical verbs (11 percent), and the distribution is also statistically significant (χ^2 (2) = 37.77, p < 0.001). The pattern displayed in Table 3 bears directly on the research question presented in the previous section and it will be the main focus of Section 6, where we will discuss its implications.

In the last step of the analysis, we investigated the effect of agreement on pronoun omission within each verb class. Table 4 displays the distribution of omitted and overt pronouns in Indian English with respect to the presence or absence of subject-verb agreement in lexical verbs and non-modal auxiliaries:[3]

Table 4: Effect of agreement on pronoun omission per verb class in Indian English

Presence of Agreement	Verb Class			
	Lexical		Non-modal Auxiliaries	
	Omitted	Overt	Omitted	Overt
Agreement	18.6% (18)	81.4% (79)	4% (15)	96% (356)
No Agreement	7.1% (71)	92.9% (930)	3.6% (32)	96.4% (846)

In the case of lexical verbs, omitted pronouns are much more frequent when the verb agrees with the subject of the clause (18.6 percent) than when it does not (7.1 percent), and this distribution is statistically significant (χ^2 (1) = 15.6, p < 0.001). With non-modal auxiliaries, on the other hand, the percentage of omission is practically the same whether there is agreement or not between the subject and the verb (4 percent versus 3.6 percent), and in this case the result of the chi-squared test of independence indicates non-significance (χ^2 (1) = 0.11, p > 0.05). Finally, Table 5 sets out the distribution of omitted and overt pronouns in Singapore English with respect to the presence or absence of subject-verb agreement within each verb class:

Table 5: Effect of agreement on pronoun omission per verb class in Singapore English

Presence of Agreement	Verb Class			
	Lexical		Non-modal Auxiliaries	
	Omitted	Overt	Omitted	Overt
Agreement	18.3% (23)	81.7% (103)	3.2% (16)	96.8% (480)
No Agreement	10.3% (122)	89.7% (1,067)	6.5% (52)	93.5% (747)

As shown in Table 5, the situation in Singapore English is once again similar to that in Indian English, but with some minor differences. With pronouns occurring in combination with lexical verbs, omission is much more frequent when the verb agrees with the subject (18.3 percent) than when there is no agreement (10.3 percent), and this difference is significant (χ^2 (1) = 7.42, $p < 0.01$). With non-modal auxiliaries, on the other hand, we find the opposite correlation: omitted pronouns are more frequent when the subject and the verb do not agree in person and number than when they agree. The difference is not great (6.5 percent with no agreement versus 3.2 percent with agreement), but it is statistically significant (χ^2 (1) = 6.63, $p < 0.05$). Overall, it seems that the frequency of omitted pronouns when there is no subject-verb agreement is higher in Singapore English than in Indian English, which could suggest that the effect of agreement is weaker in the former than in the latter. This issue will be dealt with in Section 6.

This section has presented the results of the corpus study outlined in Section 4, with an emphasis on the frequency of omitted pronouns with respect to the different levels of the independent variables: Indian English versus Singapore English, lexical verbs versus non-modal auxiliaries versus modal auxiliaries, and presence versus lack of subject-verb agreement. Firstly, it was shown that pronoun omission is more common in Singapore English than in Indian English, and this was reflected in almost all the results presented here. Secondly, omitted pronouns occurred more frequently with lexical and modal verbs than with non-modal auxiliaries in both varieties. In Indian English, the incidence of omitted pronouns was higher with lexical than with modal verbs, whereas the opposite pattern was found in Singapore English. Finally, with lexical verbs omission was more frequent in both varieties when there was subject-verb agreement than when there was not. With non-modal auxiliaries, however, there were no differences in this respect in Indian English, and in Singapore English the correlation

was in the opposite direction: the percentage of omitted pronouns was lower when there was agreement between the subject and the verb. In the next section, these results will be related to the research question in Section 4 and to the theoretical framework described in Section 2.2.

6   Discussion

The results presented in the previous section reveal that pronoun omission is not a widespread feature in either Indian or Singapore English, attaining a frequency of only 6 percent in the former and 8.9 percent in the latter. This means that, in most cases, speakers chose to use overt pronouns in order to establish anaphoric relations. This is not surprising if we take into account that English is a language with poor agreement morphology, and hence that agreement is not present in most cases to aid in the recovery of the antecedents of omitted pronouns (see Section 2.2). Pronoun omission is more frequent in Singapore than Indian English, which may be related to existing attitudes towards the local varieties in their respective speech communities. As mentioned in Section 3.1, the English-speaking population of India has an exonormative orientation; i.e., they hold standard varieties (mostly British English) as the correct norms and as the ones that should be followed. The local variety is thus considered to be a deviation from the standard, as exemplified by widespread complaints about the low quality of the English spoken in India (Mukherjee 2007). Many speakers still perceive innovative features such as pronoun omission as mistakes, which works as a conservative force preventing their use. In Singapore, the local variety, with all its particularities, is a marker of national identity. Departures from the standard varieties are not considered mistakes, and therefore there is nothing preventing their use.

In Section 4, we formulated the question of whether agreement has a facilitating effect with respect to pronoun omission, and we hypothesized, in line with previous research, that omitted pronouns are more frequent when they occur with verbs that agree with the subject. The following cline of verb classes was then established on the basis of the number of forms of the verb that unambiguously code the person and number of the subject: non-modal auxiliaries > lexical verbs > modal auxiliaries. It was expected that the percentage of omission would decrease down the cline; i.e., omitted pronouns would be more frequent with non-modal auxiliaries than with lexical verbs, and more frequent with lexical verbs than with modal auxiliaries. The results, however, showed almost the opposite distribution: omission was less frequent with non-modal auxiliaries than with lexical and modal verbs in both varieties. The effect of agreement was

also assessed within verb classes, and it was found that it does have a facilitating effect when pronouns combine with lexical verbs: the frequency of omission was higher in both varieties when there was subject-verb agreement, which shows some support for Cole's (2009, 2010) hypothesis that agreement functions as an aid in recovering the antecedent of omitted pronouns. With non-modal auxiliaries, on the other hand, agreement seemed not to have such a facilitating influence, and this suggests that there is another variable at play that blocks omission and the potential influence of agreement in this class of verbs. Our working hypothesis is that the frequency of co-occurrence inhibits pronoun omission with non-modal auxiliaries and that agreement is not strong enough in Indian and Singapore English to override this effect.

Usage-based approaches to grammar emphasize the importance of frequency in shaping linguistic structure (cf. Barlow and Kemmer 2000; Bybee and Hopper 2001). Among the effects of frequency found in previous research, there is one that bears directly on the finding that the percentage of omitted pronouns is lower with non-modal auxiliaries: the role of repetition and frequency of co-occurrence in the creation of complex linguistic units (Haiman 1994; Bybee 1998, 2001, 2002, 2006; Bybee and Thompson 1997; Krug 1998; Bybee and Scheibman 1999; Scheibman 2000). When a string of words co-occurs very frequently in discourse, it may become stored in memory as a single complex unit. Haiman (1994: 8) has called this process "chunking", and it involves processing a recurring sequence of words as a whole. The words forming the complex unit are not treated as independent items within the construction; i.e., chunking implies the progressive loss of the internal constituent structure of the unit and the blurring of the boundaries between its component parts. The individual words, however, still retain their own mental representations, which are stored in memory independently from the chunk. The automatization of the processing of strings of words results in two additional effects. Firstly, the new unit is stored in, and retrieved from, memory as a single entity, which means that it is also articulated in production as a whole. This sets off reductive processes within the boundaries of the unit that make it more efficient to process. Krug (1998) has shown that the frequency of co-occurrence of a potentially contractible string of words is a factor affecting the incidence of contraction, so that the more often two words occur together, the more times they occur in a contracted form. Pronouns, which are high-frequency elements, are usually followed by verbs in a reduced form. Another study showing that complex units become reduced is Bybee and Scheibman (1999). They argue that the auxiliary *don't* occurs in a phonologically abbreviated form when it is part of a sequence of words that are frequently produced together, such as *I don't know*. These studies

provide evidence that chunking results in reduction that makes the unit easier to articulate. The second consequence of processing and storing strings of words as single chunks is that a new unit becomes more autonomous from its component parts and may acquire new unrelated meanings (Haiman 1994; Bybee and Scheibman 1999).

MacWhinney (2001) argues that chunking comes about both through production and perception. He distinguishes between three types of chunks: "Gestalts" or "perceptual chunks," "avalanches" or "motoric chunks," and "memory chunks" (MacWhinney 2001: 459). Perceptual chunks refer to units formed by elements that are perceived as a whole. Motoric chunks are "serial strings of behaviours in which the triggering of the beginning of the string leads to the firing of all its component pieces" (MacWhinney 2001: 459), i.e., sequences of words that are produced as a single unit. Finally, memory chunks are strings of words that are stored in memory as indivisible entities. MacWhinney claims that memory chunks come into being both through perceptual and motoric chunking. First, repetition of a string of words creates a motoric chunk, which in turn gets reduced phonologically so that it becomes easier to articulate. This abbreviated new form is then perceived as a chunk and stored in memory as a whole.

With respect to pronoun omission, non-modal auxiliaries occur very frequently in combination with pronouns: the pronoun *I* followed by *be* in any of its first person singular forms appeared 183 times in the texts taken from ICE-IND and 156 times in ICE-SIN. Moreover, 46.2 percent (1,202) and 43.5 percent (1,227) of all overt pronouns were followed by a non-modal auxiliary in Indian English and in Singapore English respectively, despite the fact that this class of verbs only comprises *be*, *have* and *do*. This high co-occurrence frequency may have to do with general principles of communication and semantics (Bybee 2007). First of all, people usually talk about themselves and the human beings around them, and personal pronouns are the device most commonly used to refer to this type of antecedents. Secondly, elements that have abstract and general meanings, such as auxiliary verbs, can occur in combination with many different words, because their meaning can be applied almost to any situation. Therefore, as suggested by Bybee (2001), sequences of words made up by a pronoun plus a form of an auxiliary may be stored in memory as single units due to their high frequency of co-occurrence. This analysis is supported by the fact that non-modal auxiliaries often appear in a contracted form when they are preceded by personal pronouns (cf. Krug 1998). These units formed by a pronoun plus a non-modal auxiliary are accessed and retrieved from memory as a whole; i.e., the pronoun and the verb are no longer independent items that can be individu-

ally selected and omitted. As a result, pronoun omission is predicted to be inhibited or blocked in these cases, and this is the reason why there is such a low percentage of omitted pronouns with non-modal auxiliaries in Indian English and Singapore English. The pronoun and the verb, however, still exist as autonomous units with their own entries in memory, and they can be combined to produce the same string of words. When this happens, the pronoun can be omitted, and this is why we still find some instances of pronoun omission with non-modal auxiliaries in both varieties.

Finally, the results of the corpus study presented in Section 5 show that the relation between subject-verb agreement and pronoun omission is not the same in Indian English and in Singapore English. Agreement seems to have a stronger facilitating effect in the former than in the latter: there is a higher frequency of omission without agreement in Singapore English (10.3 percent within the class of lexical verbs and 6.5 percent within non-modal auxiliaries) than in Indian English (7.1 percent within lexical verbs and 3.6 percent within non-modal auxiliaries). Moreover, the highest percentage of omitted pronouns in Singapore English occurs with modal auxiliaries. Subject-verb agreement is attested in many of the languages that form India's linguistic substratum, and their speakers are used to exploit the presence of agreement morphology in order to identify the formal features of omitted pronouns. This results in a stronger facilitating effect of agreement in Indian English. Chinese (Singapore's main substrate language), on the other hand, is an isolating language, which means that the process whereby the referents of omitted pronouns are identified and recovered is independent from morphology. As a consequence, Singapore English speakers do not rely as much on agreement morphology, and for this reason pronoun omission is more frequent without agreement in Singapore English than in Indian English. Despite these differences, however, the relation between omission and agreement is similar in both varieties: agreement only has a minor facilitating effect within the class of lexical verbs because English is a language with poor agreement morphology that does not code the person and number of the subject on most occasions.

7   Conclusions

In the present paper, we have investigated the relation between agreement and the omission of subject pronouns in Indian English and Singapore English. Subject-verb agreement is a powerful predictor of pronoun omission cross-linguistically because it aids in the identification and retrieval of the referents of omitted pronouns. The results of the corpus study presented in Section 5, however,

showed a much more complicated picture. Contrary to expectations, omitted pronouns occurred less frequently in both varieties with the non-modal auxiliaries *be*, *have* and *do* than with lexical and modal verbs, even though the former are the class of verbs in English that have the highest number of forms that agree in person and number with the subject. Agreement was then found to have a weak facilitating effect, but only when pronouns were followed by lexical verbs. This distribution was linked to previous research on the effects of frequency of co-occurrence: it was suggested that there may be complex units formed by a pronoun plus a non-modal auxiliary that are stored in memory as a whole due to their high co-occurrence frequency, and that this would inhibit the omission of pronouns.

Despite minor differences between Indian English and Singapore English which were attributed to the status of English in their respective speech communities and to the different substrate languages of the populations, the results of the study show that subject-verb agreement does not have a strong facilitating effect in either variety: pronoun omission only increased in the presence of agreement when pronouns occurred with lexical verbs. These findings tie in with work on sentence processing that argues that there are certain cues in languages that are stronger than others, and those are the ones that users rely on when processing linguistic structures (cf. MacWhinney 2001, and references therein). English is a language with poor agreement morphology in which the verb only agrees with the subject in a limited set of contexts. Subject-verb agreement is not a strong cue in (any variety of) English, and therefore users do not rely on it as much as in other languages. Still, we found that agreement facilitated pronoun omission with lexical verbs in Indian English and Singapore English, which shows support for Cole's (2009, 2010) theory that agreement functions as an aid in recovering the referents of omitted pronouns in those languages in which it is attested. This in turn suggests that speakers make efficient use of the linguistic material available to them by taking advantage of the cues in a structure that may help them to process it faster and more easily.

We believe that the present article has provided interesting findings regarding the connection between agreement and pronoun omission in Indian English and Singapore English. Additionally, we have suggested the existence of another variable that may influence the omission of subject pronouns: the frequency of co-occurrence between the pronoun and the verb. To the best of our knowledge, this link between frequency and omission has not been observed before. In this paper, we have formulated a hypothesis based on previous work on how frequency of co-occurrence affects linguistic structure, but this requires empirical testing, which must be left for future research.

Acknowledgements

Notes

1. Even though third person singular personal pronouns in English have separate forms according to their gender, this feature is not morphologically marked in verbs and, therefore, it must always be recovered from context if the pronoun is omitted.
2. *Be* is the English verb in which more forms agree in person and number with the subject. The auxiliaries *have* and *do* agree only with third person singular present subjects, as in the case of lexical verbs, but they don't lose their agreement morphology when they occur in negative or interrogative clauses.
3. This last step of the analysis was restricted to lexical verbs and non-modal auxiliaries. Modal auxiliaries were excluded because they lack subject-verb agreement, i.e., no form of any modal verb agrees in person and/or number with the subject.

Corpora and electronic resources

ICE-IND: *International Corpus of English – the Indian Component*. 2002. Compiled by S. V. Shastri at Shivaji University, India, and Gerhard Leitner at Freie Universität Berlin. http://ice-corpora.net/ice/download.htm

ICE-SIN: *International Corpus of English – the Singaporean Component*. 2002. Compiled by Paroo Nihilani, Ni Yibin, Anne Pakir and Vincent Ooi. Singapore: The National University of Singapore. http://ice-corpora.net/ice/download.htm

R Development Core Team. 2015. *R: A language and environment for statistical computing*. Vienna: The R Foundation for Statistical Computing. http:// www.R-project.org.

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

References

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24: 65–87.

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.

Ariel, Mira. 2001. Accessibility theory: An overview. In T. Sanders, J. Schliperoord and W. Spooren (eds.). *Text representation*, 29–87. Amsterdam: John Benjamins.

Bao, Zhiming. 2001. The origins of empty categories in Singapore English. *Journal of Pidgin and Creole Languages* 16: 275–319.

Barlow, Michael and Suzanne Kemmer (eds.). 2000. *Usage-based models of language*. Stanford: Center for the Study of Language and Information.

Bhatt, Rakesh M. 2004. Indian English: Syntax. In B. Kortmann and E. Schneider *et al*. (eds.). *A handbook of varieties of English*, 1116–1130. New York: Mouton de Gruyter.

Bybee, Joan L. 1998. The emergent lexicon. In M. C. Gruber, D. Higgins, K. S. Olson and T. Wysocki (eds.). *CLS 34: The panels*, 421–435. University of Chicago: Chicago Linguistic Society.

Bybee, Joan L. 2001. Frequency effects on French liaison. In J. L. Bybee and P. J. Hopper (eds.). *Frequency and the emergence of linguistic structure*, 337–359. Amsterdam: John Benjamins.

Bybee, Joan L. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261–290.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–733.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. New York: Oxford University Press.

Bybee, Joan L. and Paul J. Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, Joan L. and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in American English. *Linguistics* 37: 575–596.

Bybee, Joan L. and Sandra A. Thompson. 1997. Three frequency effects in syntax. In M. Juge and J. Moxley (eds.). *Proceedings of the Twenty-third Annual Meeting of the Berkeley Linguistics Society, February 14–17, 1997: General session and parasession on pragmatics and grammatical structure*, 378–388. Berkeley: Berkeley Linguistics Society.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of Government and Binding*. Cambridge, MA: The MIT Press.

Cole, Melvyn D. 2009. Null subjects: A reanalysis of the data. *Linguistics* 47: 559–587.

Cole, Melvyn D. 2010. Thematic null subjects and accessibility. *Studia Linguistica* 64: 271–320.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Haegeman, Liliane. 1994. *Introduction to Government and Binding Theory*. 2nd edition. Oxford: Blackwell.

Haiman, John. 1994. Ritualization and the development of language. In W. Pagliuca (ed.). *Perspectives on grammaticalization*, 3–28. Amsterdam: John Benjamins.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15: 531–574.

Huang, Yan. 1992. Against Chomsky's typology of empty categories. *Journal of Pragmatics* 17: 1–29.

Huang, Yan. 2000. *Anaphora: A cross-linguistic study*. Oxford: Oxford University Press.

Huddleston, Rodney and Geoffrey K. Pullum *et al.* 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.

Kachru, Yamuna. 2006. *Hindi*. Amsterdam: John Benjamins.

Krug, Manfed. 1998. String frequency: A cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26: 286–320.

Lange, Claudia. 2012. *The syntax of spoken Indian English*. Amsterdam: John Benjamins.

Leimgruber, Jakob R. E. 2013. *Singapore English: Structure, variation, and usage*. Cambridge: Cambridge University Press.

Li, Charles N. and Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

MacWhinney, Brian. 2001. Emergentist approaches to language. In J. L. Bybee and P. J. Hopper (eds.). *Frequency and the emergence of linguistic structure*, 449–470. Amsterdam: John Benjamins.

Moag, Rodney F. and Robert Poletto. 1991. Discourse level evidence for South Asia as a linguistic area. *Language Sciences* 3: 229–254.

Mukherjee, Joybrato. 2007. Steady states in the evolution of New Englishes: Present-day Indian English as an equilibrium. *Journal of English Linguistics* 35: 157–187.

Neeleman, Ad and Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38: 671–716.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17: 501–558.

Ruppenhofer, Josef and Laura A. Michaelis. 2010. A constructional account of genre-based argument omissions. *Constructions and Frames* 2: 158–184.

Scheibman, Joanne. 2000. *I dunno*: A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32: 105–124.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Speas, Margaret. 1994. Null arguments in a theory of economy of projection. In E. Benedicto and J. T. Runner (eds.). *University of Massachusetts Occasional Papers in Linguistics 17: Functional projections*, 179–208. Amherst: Graduate Linguistics Student Association, University of Massachusetts.

Speas, Margaret. 2006. Economy, agreement and the representation of null arguments. In P. Ackema, P. Brandt, M. Shoorlemmer and F. Weerman (eds.). *Arguments and agreement*, 35–75. Oxford: Oxford University Press.

Taraldsen, Knut T. 1980. *On the nominative island condition, vacuous application, and the* that-*trace filter*. Bloomington, Ind.: Indiana University Linguistics Club.

Wee, Lionel. 2004. Singapore English: Morphology and syntax. In B. Kortmann and E. Schneider *et al*. (eds.). *A handbook of varieties of English*, 1116–1130. New York: Mouton de Gruyter.