A CROSS-VARIETAL STUDY OF PRONOUN OMISSION IN ENGLISH: SIMPLIFICATION AND SUBSTRATE EFFECTS

Iván Tamaredo
Universidade de Santiago de Compostela
ivan.tamaredo@usc.es

Abstract

This contribution presents a study of pronoun omission in English and assesses the importance of two explanatory factors, namely simplification due to L2 acquisition and use and substrate influence, to account for this grammatical feature. The results suggest that substrate effects are of secondary importance in comparison to simplification as a predictor of pronoun omission in English, but that both factors are necessary to provide a full description of this phenomenon.

Keywords: Pronoun omission, varieties of English, simplification, substrate influence.

1. INTRODUCTION

Pronoun omission, that is, the absence of a pronominal element in the clause, is a phenomenon that has recently caught the attention of several researchers in the field of New Englishes (cf. Bao 2001; Schröter and Kortmann 2016). The present paper further explores this issue by conducting a cross-varietal survey of pronoun omission in English and assessing the explanatory power of two factors that are commonly resorted to in order to account for differences between varieties of English, namely simplification due to L2 acquisition and use and influence from the substrate languages.

The rest of the contribution is structured as follows. Sections 2 and 3 provide a brief literature review on the phenomenon of pronoun omission and the explanatory factors simplification and substrate influence, respectively. Section 4 deals with the data and the methodology of the study, followed, in section 5, by a summary and discussion of the results. Finally, section 6 presents some concluding remarks.

2. PRONOUN OMISSION

Pronoun omission refers to those cases in which there is a gap in a clause that could be filled by an overt pronoun, usually a personal pronoun. In generative approaches to grammar, the gap was claimed to be filled by an empty element with a function from the point of view of syntax but no overt phonetic realization, and researchers mainly focused on the syntactic contexts which licensed the occurrence of these elements, such as the presence of rich subject-verb agreement (Rizzi 1986). However, despite the formulation of syntactic restrictions, a complete account of this phenomenon requires considering issues of a cognitive nature. For instance, even in languages with rich subject-verb agreement, it is ultimately the presence of a highly accessible antecedent in the co-text/context which allows the omission of the subject (Cole 2010).

Present-day standard English is considered a non-pro-drop language, that is, one in which omitted pronouns are not allowed by the grammar, but they do occur in certain contexts. Instances of pronoun omission can be found in casual style in subject position at the beginning of main clauses, and in coordinate clauses, if the omitted pronoun is the subject of the second conjunct and if it is coreferential with that of the first one (Huddleston and Pullum et al. 2002, 1540-1541). However, in non-standard varieties pronoun omission is not so restricted, as attested in the *Electronic World Atlas of Varieties of English* (*eWAVE*; Kortmann and Lunkenheimmer 2013), a database containing information about 235 morphosyntactic features in 76 varieties of English. *eWAVE* includes five pro-drop features, exemplified in (1)-(5):

(1) F42: Object pronoun drop (attested in 37% of the varieties in *eWAVE*)
I told him to get some **pears**$_i$ but he forgot to bring **Ø**$_i$.

(2) F43: Subject referential pronoun drop (51%)
So you worked in the fields? Oh yes all the time. **Ø** Picked potatoes.

(3) F44: Subject dummy pronoun drop (36%)
Ø Must be getting late.

(4) F46: Deletion of *it* in referential *it is*-constructions (39%)
My grandfather and grandmother speaks **it**$_i$ and that's how I pick **it**$_i$ up and **Ø**$_i$ is still in me.

(5) F47: Deletion of *it* in non-referential *it is*-constructions (28%)
Here **Ø** is not allowed to stop the car.

3. SIMPLIFICATION AND SUBSTRATE INFLUENCE

A question that remains to be properly answered is why we find pronoun omission features in some varieties of English but not in others. Most varieties included in *eWAVE* have a history of contact between English and other languages,

and, whenever there is language contact, at least two factors must be examined: simplification due to universal mechanisms of L2 acquisition and use, and influence from substrate languages.

Deciding what counts as a simplifying grammatical feature is a complex issue since there is yet no standard metric of complexity. However, two general principles have figured prominently in the specialized literature: *economy* and *transparency*. Economy (cf. Kusters 2003, 22-25) can be defined in terms of the amount of linguistic material that is obligatorily coded in a language. It increases if fewer grammatical elements receive overt marking and decreases if categories that other languages leave unmarked are coded. Transparency (cf. Kusters 2003, 26-30) refers to the relation between meanings and their formal marking. Transparency increases when one grammatical meaning is coded by one form and decreases when there is no one-to-one relation between meanings and forms. Pronoun omission can be characterized as a simplifying feature, since it implies an increase in economy without a decrease in transparency if the omitted pronoun refers to an antecedent that is highly accessible, an intuition that was already captured in earlier studies (cf. Williams 1989).

Substrate effects can be described as the "reinforcing influence of the mother tongue on the second language, resulting in language change" (Davydova 2011, 12). Pronoun omission in varieties of English has traditionally been analysed as the result of transfer from substrate languages (cf. Schneider 2013). However, simplification and substrate influence are highly intertwined processes, and distinguishing their effects is not always an easy task (Davydova 2011, 21): in many cases, the features that are transferred can also be characterized as simple. For instance, the use of a preverbal negator *no* (F160), as in example (6) from *eWAVE*, is characteristic of pidgins and creoles.

(6) Me **no** know (Trinidadian Creole)

Many of the substrate languages of English-based pidgins and creoles have an invariant preverbal negator marker (cf. Schneider 2012, 888), which suggests that its presence in these varieties is due to substrate influence. On the other hand, Szmrecsanyi and Kortmann (2009, 70) argue that the preverbal negator *no* eases L2 acquisition and use, as evidenced by its frequency in learner speech. This example shows that discriminating between simplification and substrate effects is not always straightforward.

4. DATA AND METHODOLOGY

The main goal of the present paper is to measure the importance of simplification and substrate influence as explanatory factors for the occurrence of pronoun omission in varieties of English. To this purpose, a cross-varietal study

was conducted using *eWAVE* as a source of data, which provides information on which features are attested in each of the varieties included and how frequently they occur by means of the following values:

– A: the feature is pervasive or obligatory
– B: the feature is neither pervasive nor rare
– C: the feature exists, but is rare
– D: the feature is absent from the grammar
– X: the feature is not applicable
– ?: no information on the feature is available

Using the values for the five pronoun omission features mentioned in section 2, two indexes were computed. The first one, labelled *pro-drop attestation*, measures if pronoun omission is present in the variety. It is a categorical variable with two levels: attested if at least one of the features is given an A, B, or C value in *eWAVE*, and not attested otherwise. The second index, *pro-drop pervasiveness*, measures how frequent pronoun omission is in those varieties in which it is attested. It is a numeric variable that is calculated by transforming A, B, and C values into numbers (1, 0.6, and 0.3 respectively) and then summing them to obtain a global pervasiveness score for each variety. As an illustrative example, consider Tables 1 and 2, which contain the pro-drop attestation and pervasiveness indexes for White Zimbabwean (WhZimE) and White South African English (WSAfE):

TABLE 1: ATTESTATION AND PERVASIVENESS IN WHITE ZIMBABWEAN ENGLISH

WhZimE	Pro-drop attestation	Pro-drop pervasiveness
Object drop	D	-
Subject drop (referential)	?	-
Subject drop (dummy)	D	-
Deletion of referential *it*	D	-
Deletion of non-ref. *it*	D	-
Pronoun omission	Not Attested	-

TABLE 2: ATTESTATION AND PERVASIVENESS IN WHITE SOUTH AFRICAN ENGLISH

WhSAfE	Pro-drop attestation	Pro-drop pervasiveness
Object drop	B	0.6
Subject drop (referential)	B	0.6
Subject drop (dummy)	?	-
Deletion of referential it	D	-
Deletion of non-ref. it	D	-
Pronoun omission	Attested	1.2

In WhZimE none of the features have an A, B, or C value, so pronoun omission is not attested and we do not calculate its pervasiveness. In WhSAfE, on the other hand, both object drop and subject drop (referential pronouns) receive B values, which means that pronoun omission is attested in this variety and we can calculate its frequency by transforming the B values into numbers (0.6) and summing them, resulting in a global pervasiveness of 1.2.

Finally, the varieties were classified according to their *type*, functioning as a proxy for L2 simplification effects, and the *geographical area* of the world in which they are spoken, functioning as a proxy for substrate effects (Kortmann 2013). To this purpose, the classification provided in *eWAVE* was followed, which distinguishes between eight regions (British Isles, America, Africa, Asia, Australia, Pacific, Caribbean, and South Atlantic) and five variety types (Traditional low-contact L1s, High-contact L1s, Indigenized L2s, Pidgins, and Creoles). However, the High-contact L1 category had to be further refined, in line with Szmrecsanyi (2012), since it includes both L1 varieties that were or are still affected by contact with other languages and varieties that are the result of dialect contact. Since in the present study it is assumed that pronoun omission is the result of simplification processes due to universal mechanisms of L2 acquisition and use, these two different High-contact L1 categories had to be separated, resulting in six variety types: Traditional L1s, Dialect-contact L1s,[1] Shift L1s,[2] Indigenized L2s, Pidgins, and Creoles.

5. RESULTS AND DISCUSSION

5.1. Attestation

The first set of results reported on in this section deals with pro-drop attestation, which, as mentioned in section 4, measures whether pronoun omission is attested or not in a variety. Variety type is the only statistically significant factor, as confirmed by the results of Pearson's Chi-squared test ($\chi^2 = 13.38$, d.f. $= 5$, $p < 0.05$), whereas area does not have a significant effect. Table 3 shows the attestation rates for the six variety types.

[1] Dialect-contact L1s result from the convergence of different dialects of English in the same speech community.

[2] Shift L1s are varieties that were mostly used as L2s in the past but that have been adopted as L1s by the majority of their speech community.

TABLE 3: PRO-DROP ATTESTATION AS FUNCTION OF VARIETY TYPE		
Variety type	Pro-drop attested	Pro-drop not attested
Traditional L1	40% (4)	60% (6)
Dialect-contact L1	37.5% (3)	62.5% (5)
Shift L1	85.7% (12)	14.3% (2)
Indigenized L2	88.9% (16)	11.1% (2)
Pidgin	71.4% (5)	28.6% (2)
Creole	73.7% (14)	26.3% (5)

Pronoun omission is much more frequently attested in those varieties which are/were affected by L2 acquisition and use, that is, Shift L1s, Indigenized L2s, Pidgins, and Creoles: their attestation rates range from 71.4% to 88.9%. On the other hand, Traditional and Dialect-contact L1s exhibit a much lower percentage of attestation (40% and 37.5%, respectively). This distribution clearly shows that L2 varieties are more prone to the omission of pronouns than Traditional and Dialect-contact L1s, which suggests that simplification due to L2 acquisition and use is indeed an important explanatory factor for the occurrence of pronoun omission in varieties of English.

This conclusion is further reinforced by random rorest and conditional inference tree analyses, with pro-drop attestation as dependent variable and variety type and area as predictors.[3] Figure 1 plots the explanatory power of the two predictors, extracted from the results of a random forest analysis, and Figure 2 plots the statistically significant split in the data resulting from a conditional inference tree.

What these two figures show is that, when tested together for their conjoined effect, only variety type emerges as an important explanatory factor of the degree of attestation of pronoun omission in varieties of English, whereas area plays no role at all. Furthermore, as can be seen in Figure 2, the only significant division is that between Traditional and Dialect-Contact L1s (with a predicted

[3] Conditional inference trees predict the probability of the values of a dependent variable by means of a series of binary splits in the data. The test takes the predictor that is more strongly correlated with the dependent variable, for instance, variety type, and considers whether dividing the data in two, for instance, with Traditional and Dialect-contact L1s in one branch of the tree and the remaining types in the other, creates two new subsets of cases in which one of the values of the dependent variable is more likely. This process is then repeated for each new partition until no further significant splits are found. Random forests are computed from many trees and assess how important each predictor is to account for the variation found in the dependent variable (cf. Tagliamonte and Baayen 2012).

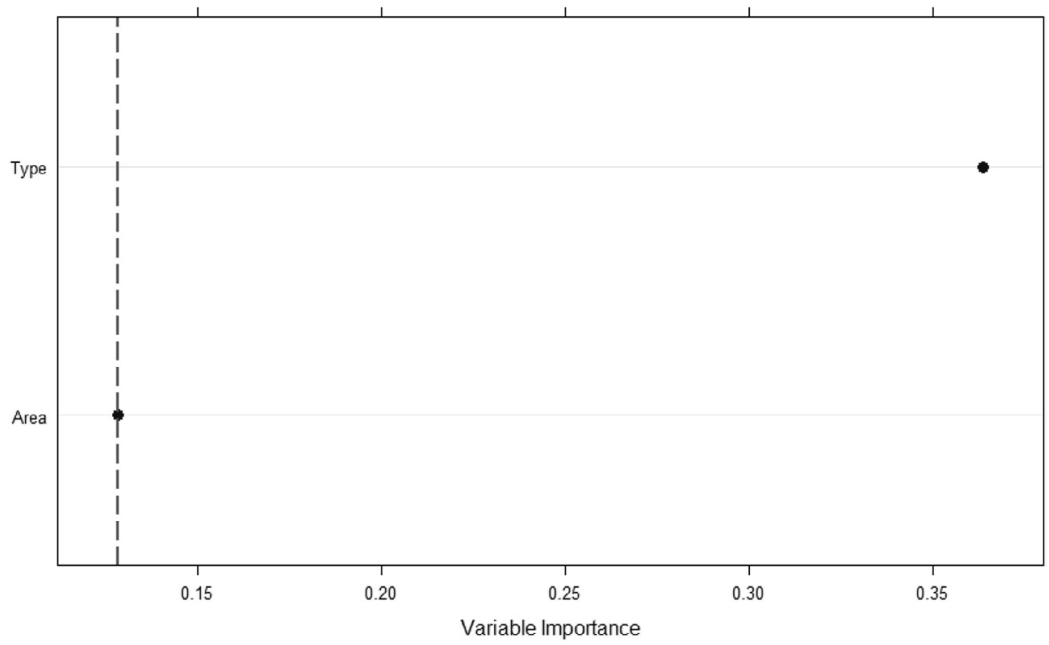

Figure 1: Relative importance of variety type and area as predictors of pro-drop attestation.

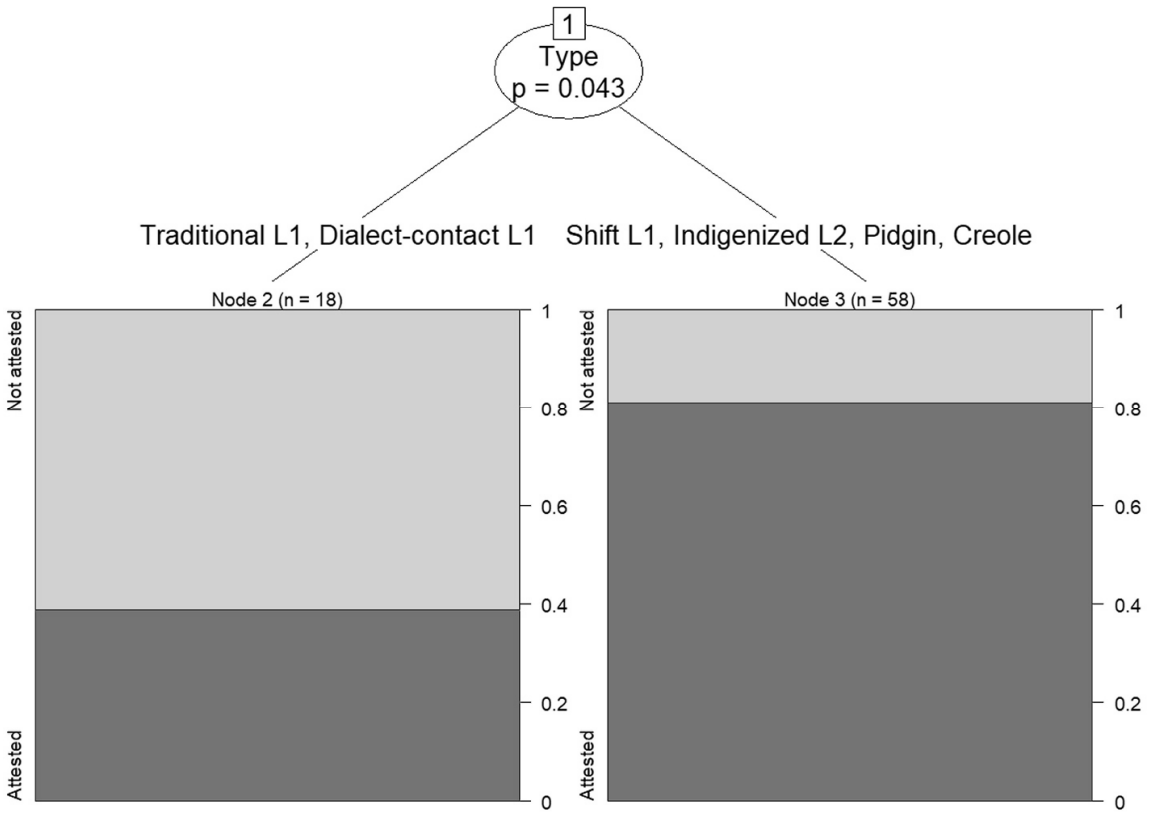

Figure 2: Conditional inference tree with variety type and area as predictors of pro-drop attestation.

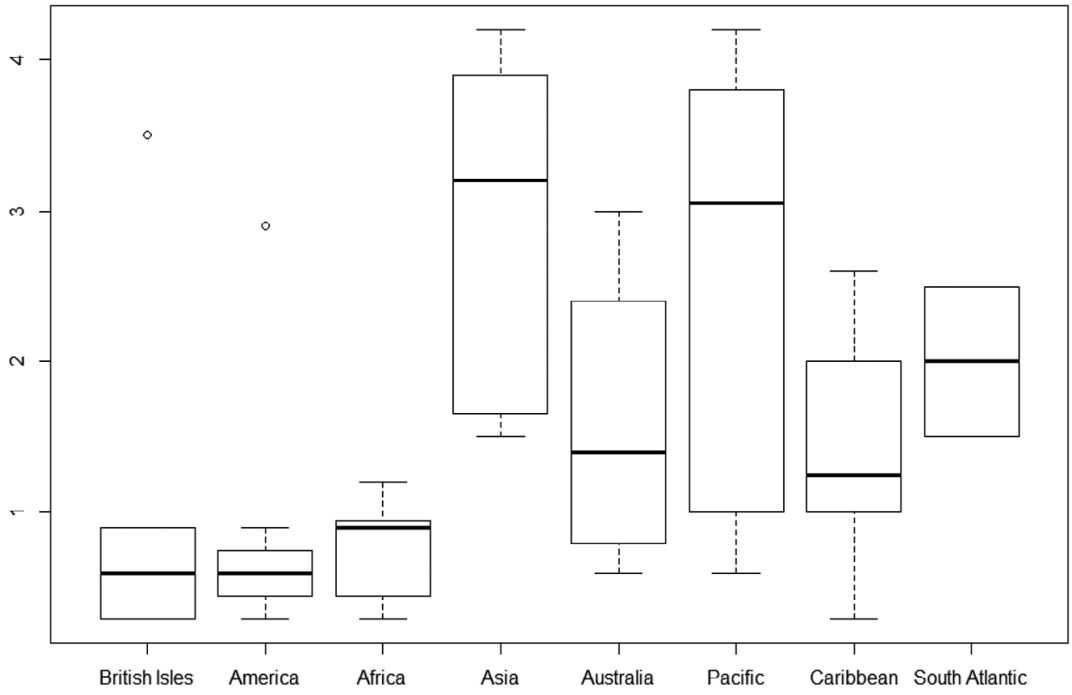

Figure 3: Median pro-drop pervasiveness per geographical area.

probability of pro-drop attestation of almost 40%), on the one hand, and Shift L1s, Indigenized L2s, Pidgins, and Creoles (with a probability of approximately 80%), on the other. This finding suggests that it is not the specific variety type what influences the attestation of pronoun omission features, but rather whether the varieties are/were affected by L2 acquisition and use.

5.2. Pervasiveness

Pro-drop pervasiveness quantifies how frequently pronoun omission occurs in those varieties in which it is attested. In this case, only area is significant, as the results of Kruskal-Wallis rank sum test ascertain ($\chi^2 = 25.07$, d.f. = 7, $p < 0.001$), while variety type does not have a significant effect. Table 4 and Figure 3 show the median pervasiveness values for the eight regions.

TABLE 4: MEDIAN PRO-DROP PERVASIVENESS PER GEOGRAPHICAL AREA

Area	Pro-drop pervasiveness
British Isles	0.6
America	0.6
Africa	0.9

continue →

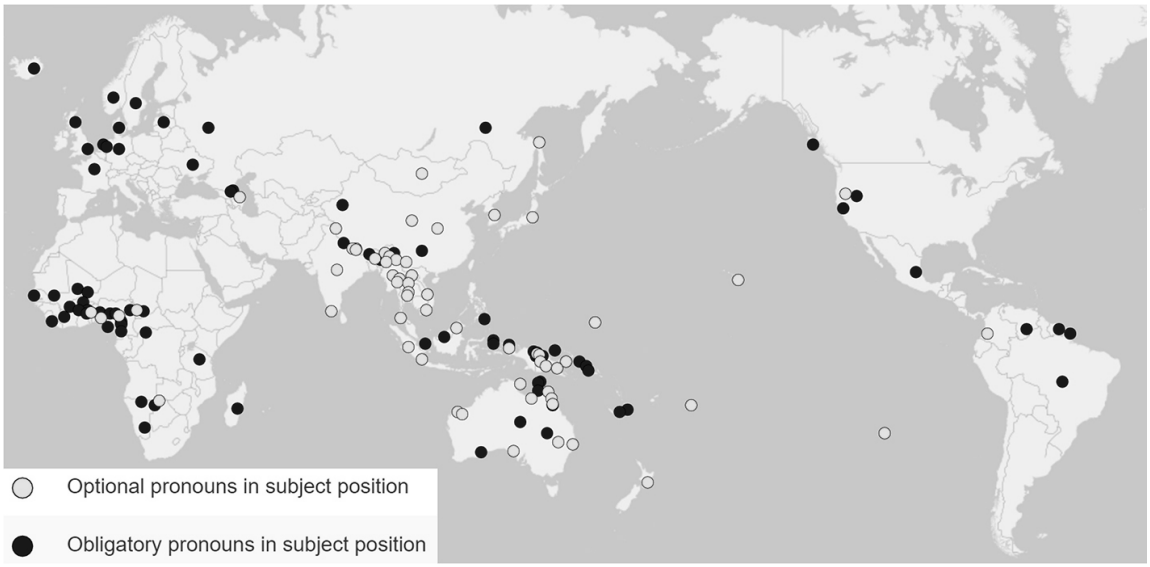

Figure 4: Expression of pronominal subjects in *WALS*.

TABLE 4: MEDIAN PRO-DROP PERVASIVENESS PER GEOGRAPHICAL AREA

Area	Pro-drop pervasiveness
Asia	3.2
Australia	1.4
Pacific	3.05
Caribbean	1.25
South Atlantic	2

Pronoun omission seems to be particularly pervasive in Asia and the Pacific, while it occurs much less frequently in the varieties located in the British Isles, America, and Africa, with Australia and the Caribbean occupying and intermediate position. As mentioned in section 4, area is used in the present paper as a proxy for substrate influence, that is, it is assumed that the high pro-drop pervasiveness exhibited by, for instance, Asian varieties is due to the reinforcing influence of their substrate languages. This claim can be substantiated by examining the *World Atlas of Language Structures Online* (Dryer and Haspelmath 2013), a database mapping 192 grammatical features in up to 2,679 different languages. One of these features provides information on how several languages code pronominal subjects, and, as shown in Figure 4, the resulting distribution closely resembles the one found for the varieties of English in *eWAVE*: while languages with optional expression of pronominal subjects are mostly located in Asia, the Pacific, and Australia, the three areas in *eWAVE* with the highest median pro-drop pervasiveness values, those with obligatory subject pronouns can be found in Africa, America, and the British Isles, the areas with the lowest values.

6. CONCLUSIONS

The aim of the present contribution was to measure the explanatory power of two factors, simplification due to L2 acquisition and use and substrate influence, with respect to pronoun omission and to tease apart their independent effects. To this purpose, two indexes, pro-drop attestation and pro-drop pervasiveness, were computed on the basis of the feature values provided in *eWAVE*, and the varieties in this database were classified according to their type and the geographical area to which they belonged. The results of the study showed that type was the only variable that could predict the presence or absence of pronoun omission features in a variety, with those varieties affected by L2 acquisition/use exhibiting a higher percentage attestation of these features than Traditional and Dialect-contact L1s. Area, on the other hand, accounted for pro-drop pervasiveness, that is, how frequently pronoun omission occurred in those varieties in which it was attested, with varieties spoken in Asia, the Pacific, and Australia, in which most pro-drop languages are found, displaying median higher pervasiveness values than the rest. These findings suggest that substrate influence is only of secondary importance in comparison to simplification, but both factors are needed if we want to reach a fuller understanding of the linguistic phenomenon of pronoun omission in English.

REFERENCES

Bao, Zhiming. 2001. "The Origins of Empty Categories in Singapore English." *Journal of Pidgin and Creole Languages* 16 (2): 275-319.

Cole, Melvyn. 2010. "Thematic Null Subjects and Accessibility." *Studia Linguistica* 64 (3): 271-320.

Davydova, Julia. 2011. *The Present Perfect in Non-Native Englishes: A Corpus-Based Study of Variation*. Berlin: De Gruyter Mouton.

Dryer, Matthew, and Martin Haspelmath, eds. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [Accessed online on May 19, 2017].

Huddleston, Rodney, and Geoffrey Pullum et al. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge UP.

Kortmann, Bernd. 2013. "How Powerful Is Geography as an Explanatory Factor of Variation? Areal Features in the Anglophone World." In *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives*, edited by Peter Auer, Martin Hilpert, Anja Stukenbrock and Benedikt Szmrecsanyi, 165-194. Berlin and Boston: De Gruyter.

Kortmann, Bernd, and Kerstin Lunkenheimer, eds. 2013. *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [Accessed online on May 19, 2017].

Kusters, Wouter. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.

Rizzi, Luigi. 1986. "Null Objects in Italian and the Theory of *pro*." *Linguistic Inquiry* 17 (3): 501-558.

Schneider, Agnes. 2012. "Typological Profile: Pidgins and Creoles." In *The Mouton Atlas of Variation in English*, edited by Bernd Kortmann and Kerstin Lunkenheimer, 875-904. Berlin: De Gruyter.

Schneider, Edgar. 2013. "English as a Contact Language: The 'New Englishes'." In *English as a Contact Language*, edited by Daniel Schreier and Marianne Hundt, 131-148. Cambridge: Cambridge UP.

Schröter, Verena, and Bernd Kortmann. 2016. "Pronoun Deletion in Hong Kong English and Colloquial Singaporean English." *World Englishes* 35 (2): 221-241.

Szmrecsanyi, Benedikt. 2012. "Typological Profile: L1 Varieties." In *The Mouton Atlas of Variation in English*, edited by Bernd Kortmann and Kerstin Lunkenheimer, 827-842. Berlin: De Gruyter.

Szmrecsanyi, Benedikt, and Bernd Kortmann. 2009. "Between Simplification and Complexification: Non-Standard Varieties of English around the World." In *Language Complexity as an Evolving Variable*, edited by Geoffrey Sampson, David Gil, and Peter Trudgill, 64-79. Oxford: Oxford UP.

Tagliamonte, Sali, and Harald Baayen. 2012. "Models, Forests and Trees of York English: 'Was/were' Variation as a Case Study for Statistical Practice." *Language Variation and Change* 24 (2): 135-178.

Williams, Jessica. 1989. "Pronoun Copies, Pronominal Anaphora and Zero Anaphora in Second Language Production." In *Variation in Second Language Acquisition*, edited by Susan Gass, Carolyn Madden, Dennis Preston and Larry Selinker, 153-189. Clevedon: Multilingual Matters.