Exploring the Patterns of Verbal Agreement of Collective Noun-Headed Subjects in a Parsed Corpus of Present-Day English[*]

Yolanda Fernández-Pena
University of Vigo, Vigo, Spain
yolanda.fernandez@uvigo.es

Abstract

This study explores verbal agreement variation with collective noun-headed subjects taking plural *of*-dependents in a set of data retrieved from a parsed version of the corpus of Global Web-based English (GloWbE). The results show how syntactic distance between the subject and the verb (and also the complexity of the modifier(s) of the *of*-PP) reduces the influence that these plural *of*-PPs exert on the verb, thus lowering the rate of plural agreement.

1 Introduction

In this paper, I will first outline some basic concepts pertaining to verbal agreement with collective nouns in Section 2, and the main aims of this study as well as the methodological procedure applied in Section 3. In Section 4, I will present the data analysis, which covers two questions: regional variation and syntactic complexity. Finally, I will put forward the conclusions in Section 5.

2 Variation in Verbal Agreement with Collective Nouns

Collective noun-headed subjects, i.e. those which denote "a group of animates or inanimates" (e.g. *committee*, *party*, *series*, *number*), usually take singular or plural verbs according to whether the speaker focuses on the collectivity (1) or on its individuals (2) (Dekeyser 1975, 35fn1, 37; see also Biber et al. 1999: 188).

---

1. the *crowd* of curious onlookers *was* so thick that the paramedics had trouble lifting the body from the ambulance. [GloWbE: w_tz_g03]

2. The *crowd* of delegates *are eating* all this up [GloWbE: w_gb_g18]

The plural *of*-dependents that often accompany some collective nouns such as *crowd* in (1) and (2) usually involve further complications for subject-verb agreement. In previous literature, these plural *of*-PPs have not been devoted sufficient attention and their role in agreement has been oversimplified and attributed to mere attraction effects (cf. Levin, 2001). However, as recent corpus-based studies have attested (see Fernández-Pena 2015, 2017), and as will also be shown in this paper, these *of*-dependents significantly interact with verbal agreement in Present-Day British and American English inasmuch as they reinforce the conceptual plurality inherent to the collective noun and thus favour a higher likelihood of plural verb number, as in (2) above. This preference for the plural verb number is especially noticeable in local syntactic domains but loses significance with increasing syntactic distance between the collective subject and the verb (see Section 4.2).

3   Aims and Methodological Procedure

Taking as a point of departure prior corpus-based studies on Present-Day English (Fernández-Pena, 2015 and 2017, using data from the British National Corpus or BNC and the Corpus of Contemporary American English or COCA), the research reported here aims at broadening the scope of such investigation by exploring verbal agreement with collective noun-based subjects in the so-called 'New Englishes', a topic with few antecedents in the literature (Hundt, 2006, 2009; Jantos 2010). This study thus seeks to find further data for British and American English as well as for other regional varieties so as to corroborate/refute my previous observations as regards syntactic complexity and to provide a wider picture on these constructions.

To this end, I have examined verbal agreement with complex collective-headed constructions (N_{COLL}-*of*-N_{PL} subjects) comprising:

a. a singular collective noun (23 different collective nouns – Biber et al. 1999: 249 and Huddleston and Pullum et al. 2002: 503 – and up to 3,000 hits per collective were considered),

b. a postmodifying *of*-PP with a plural noun,

c. an inflected verbal form.

The data were retrieved from a parsed version of GloWbE, a corpus comprising twenty different varieties of English, and using the syntax query of the Dependency Bank interface of the University of Zurich.[†] The parser[‡] identified over 32,000 examples of N_{COLL}-*of*-N_{PL} subjects (reliability of 87%) which, after the manual pruning, resulted in 8,742 valid instances.

4   Data Analysis

In this section, I will present the results of the corpus-based study. First, I will comment on the issue of regional variation. The remainder of the section will be devoted to discuss how syntactic complexity affects verbal agreement in the constructions under study.

[†] For further information: http://www.es.uzh.ch/services/corpling.html

[‡] The Pro3Gres Parser: see https://files.ifi.uzh.ch/cl/gschneid/parser/

4.1 Regional Variation

In the data obtained from GloWbE we observe that *of*-PPs still exert some influence on verb number, but the results differ significantly from those in BNC and COCA, where plural agreement was considerably higher (53.28% vs. 65.42%, respectively; $\chi^2(1)$=181.4, p<.0001).
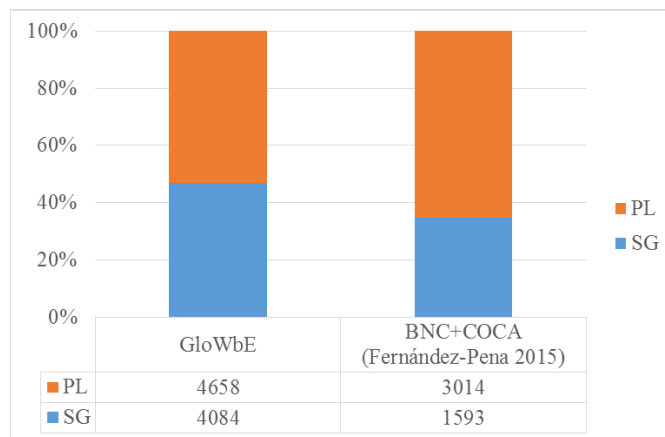

Figure 1. Verbal agreement in GloWbE vs. BNC+COCA

As commented earlier, GloWbE comprises twenty different regional varieties of English. Their heterogeneity and the potential influence of substrate (native) languages other than English are expected to be influencing these results. In fact, statistical significance is obtained for the variable 'regional variety' in the results obtained (ANOVA: $\chi^2(19)$=99.47, p<.0001). The preliminary analysis presented in Table 1 already reveals the substantial variation of the patterns of agreement of these varieties.

Regional Variety	Singular verb	Plural verb	Total
Jamaica	59 (33.15%)	119 (66.85%)	178
Nigeria	59 (35.98%)	105 (64.02%)	164
Ireland	178 (36.03%)	316 (63.97%)	494
Tanzania	71 (39.44%)	109 (60.56%)	180
Great Britain	795 (42.74%)	1,065 (57.26%)	1,860
Sri Lanka	106 (43.98%)	135 (56.02%)	241
Kenya	103 (44.59%)	128 (55.41%)	231
Pakistan	103 (44.59%)	128 (55.41%)	231
Ghana	76 (44.71%)	94 (55.29%)	170
Canada	291 (47.63%)	320 (52.37%)	611
Australia	228 (48.10%)	246 (51.90%)	474
South Africa	124 (48.63%)	131 (51.37%)	255
New Zealand	206 (49.16%)	213 (50.84%)	419
Singapore	92 (49.73%)	93 (50.27%)	185
Malaysia	85 (50.60%)	83 (49.40%)	168
India	243 (50.84%)	235 (49.16%)	478
United States	911 (52.18%)	835 (47.82%)	1,746
Bangladesh	133 (52.78%)	119 (47.22%)	252
Philippines	89 (53.29%)	78 (46.71%)	167
Hong Kong	132 (55.46%)	106 (44.54%)	238

Table 1. Verbal agreement in the twenty varieties of English in GloWbE

Table 1 evinces remarkable differences among the varieties, with only nine favouring plural agreement in more than 55% of the cases while the rest have proportions lower than 53% or even show a preference for singular verb number. Pending a more fine-grained analysis, it seems quite evident that this variability can be responsible for the general rate of agreement in the corpus as a whole.

Given these data and for the sake of simplicity, in the current section I will restrict the scope of the study to only the inner (i.e. native, see Kachru 1985) varieties of the corpus: Irish, British, Canadian, Australian, New Zealand and American English. This limitation allows us to avoid any possible interference of substrate languages in the results provided, yet it does not guarantee a more homogeneous panorama. In fact, as already illustrated in Table 1 and repeated in Table 2 for the sake of clarity, even within this set of six varieties we still observe significant variation.

Regional Variety	Singular verb	Plural verb	Total
Ireland	178 (36.03%)	316 (63.97%)	494
Great Britain	795 (42.74%)	1,065 (57.26%)	1,860
Canada	291 (47.63%)	320 (52.37%)	611
Australia	228 (48.10%)	246 (51.90%)	474
New Zealand	206 (49.16%)	213 (50.84%)	419
United States	911 (52.18%)	835 (47.82%)	1,746
TOTAL	2,609	2,995	5,604

Table 2. Verbal agreement in the inner varieties in GloWbE

As presented in Table 2, these varieties show three different trends: Irish and British English favouring plural agreement, the US showing the opposite tendency and, finally, the other three varieties with quite even figures for both numbers. This (to a certain extent) geographical classification is statistically confirmed after carrying out an association plot.
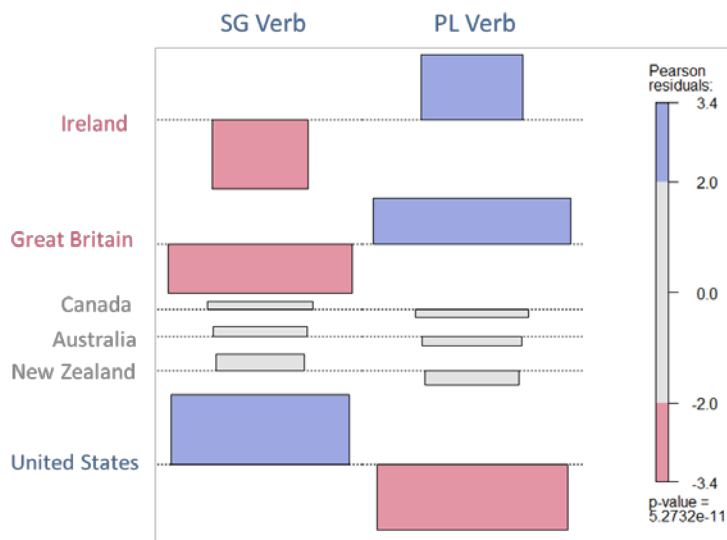

Figure 2. Pearson residuals in the inner varieties in GloWbE

As Figure 2 evinces, only the preference for plural agreement in Ireland and Great Britain and the figures for singular agreement in the US reach significance, those of Australia, New Zealand and Canada are not relevant in this regard.

Concerning the varieties of English considered in Fernández-Pena (2015, 2017), i.e. British and American English, it must be noted that very few statistically significant differences were found between the figures from BNC and those from COCA. These observations on regional variation were unexpected, as prior literature had largely attested the two aforementioned marked and divergent tendencies observed in the data from GloWbE (see Bauer 1994; Levin 2001; Depraetere 2003; Algeo 2006; or Trudgill and Hannah 2008).
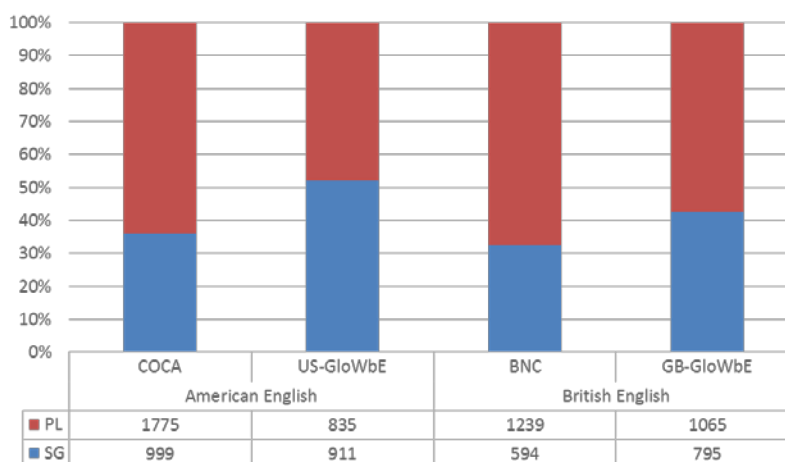

	American English		British English	
	COCA	US-GloWbE	BNC	GB-GloWbE
PL	1775	835	1239	1065
SG	999	911	594	795

Figure 3. Verbal agreement in American and British English in COCA, BNC and GloWbE

Figure 3 evinces the aforementioned differences. Hence, in both the American and the British component of GloWbE we observe significantly lower proportions of plural agreement than those in COCA and BNC (i.e. 47.82% and 57.26% in GloWbE vs. 63.99% in COCA and 67.59%; $\chi^2(1)$, p<.0001). Pending further examination, the evidence seems to suggest that such divergence may stem from the very different genre and registers of these three corpora: BNC and COCA comprising more formal texts, ranging from academic texts to fiction, popular magazines or newspapers,[§] whereas GloWbE contains only blogs and webpages.

4.2   Syntactic Complexity

As already commented, the patterns of verbal agreement of collective nouns tend to vary according to the holistic or individualistic (i.e. singular vs. plural) conceptualisation of the group. Earlier investigations on this topic have found, however, that only the latter prevails over distance. In fact, in his study Levin (2001: 93) attests that

[t]he probability of the [verbal] target agreeing with the syntactic form of the controller decreases with distance in words [since] distance decreases the focus on the grammatical number of the controllers, and the referents of the controller become more prominent.

In other words, and drawing from Corbett's (1979) Agreement Hierarchy, Levin's (2001) investigation explains that syntactic boundaries constrain our short-term memory to the extent that the formal characteristics of the collective noun are less likely to be activated over distance. As a result, only the semantic properties (i.e. the plurality) of the collective prevail and eventually trigger higher proportions of plural verb number. The N_{COLL}-of-N_{PL} constructions under scrutiny here comprise both a semantically plural collective noun and a morphologically and semantically plural noun within the *of*-PP. Contrary to expectations, Fernández-Pena (2015, 2017) demonstrated that N_{COLL}-of-N_{PL} subjects do

---

[§] Only the written components of BNC and COCA have been considered.

not take a higher rate of plural verbal patterns but that, by contrast, increasing syntactic distance and complexity significantly lowers their proportions of plural verb number.

The data from GloWbE further corroborate my prior observations for the four syntactic configurations of the *of*-PP examined here:

3. *of* + bare NP: your little band [of fanatics] are looking smaller and smaller [GloWbE: w_nz_b10]

4. *of* + *premodifier* + NP: a gang [of *armed* men] are reported to have raped five women [GloWbE: w_jm_b04]

5. *of* + bare NP + *postmodifier*: A group [of students *from Jamaica College*] have declared that adults are failing their children [GloWbE: w_jm_g10]

6. *of* + *premodifier* + NP + *postmodifier*: a large number [of *spiritual* children *in this country*] are concerned about their future [GloWbE: w_pk_g06]

Figure 4 and Figure 5 below present the percentages of agreement for these four types of constituent structure of the *of*-PP in American and British English in GloWbE, respectively. In both cases, when agreement is established locally, i.e. with no intervening material between the subject and the verb, as in (3) and (4) above, plural verb number is significantly high: 47.94% and 50.30% in American English vs. 55.57% and 65.36% in British English.
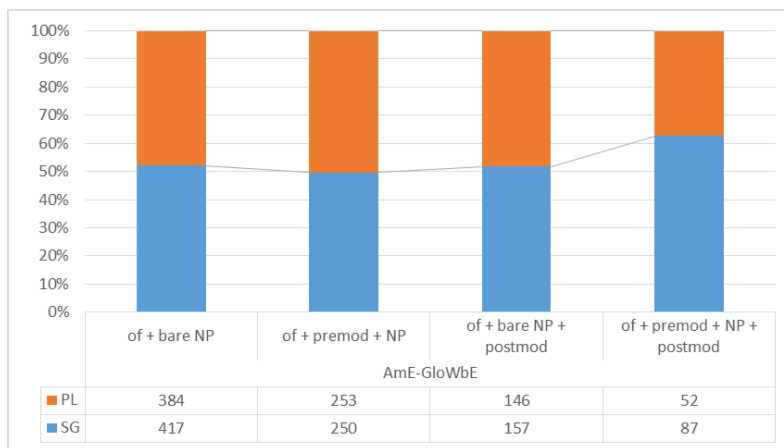

	of + bare NP	of + premod + NP	of + bare NP + postmod	of + premod + NP + postmod
		AmE-GloWbE		
■ PL	384	253	146	52
■ SG	417	250	157	87

Figure 4. Verbal agreement in bare, premodified and/or postmodified *of*-dependents in American English in GloWbE

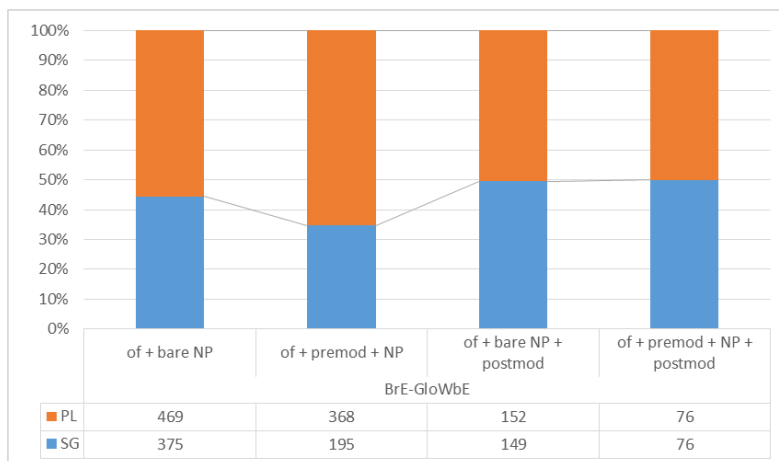|	of + bare NP	of + premod + NP	of + bare NP + postmod	of + premod + NP + postmod
			BrE-GloWbE	
PL	469	368	152	76
SG	375	195	149	76

Figure 5. Verbal agreement in bare, premodified and/or postmodified *of*-dependents in British English in GloWbE

As postmodifiers enter the picture, and thus syntactic distance between the subject and the verb increases (see (5) and (6) above), plural verb number experiments a decrease in both varieties of English. In fact, postmodified structures have a lower rate of plural agreement than non-postmodified structures (i.e. 48.18% and 50.50% in American and British English, respectively), a decrease which is even more noticeable in the most complex structure: *of* + *premodifier* + NP + *postmodifier* (i.e. 37.41% in American English – $\chi^2(1)=4.87$, p<.0001 – vs. 50.00% in British English).

In order to determine further factors influencing this decrease of plural verb number over distance, I have also investigated the typology of the postnominal modifiers (PNMs). These *of*-PPs are most frequently modified by another prepositional phrase (about 47%) or a non-finite clause (i.e. TP; 21%), with relative clauses and noun phrases showing a lower incidence (about 16% and 7%, respectively). As regards their constituent structure, these PNMs can be classified according to their non-clausal – (7) and (8) – or clausal – (9) and (10) – status.

7. PP: the majority of people *in this country* have finally accepted [GloWbE: w_gb_g14]

8. NP: A band of male lions, *often brothers*, have a range which they mark patrol and defend. [GloWbE: w_tz_g10]

9. TP: The vast majority of kiwis *living in Australia* don't have enough points to qualify for PR [GloWbE: w_nz_g07]

10. REL CL: a group of connected rooms, *which are built as living quarters*, have a separate entrance [GLoWbE:w_gh_g07]

It must be noted that collective-headed subjects containing clausal PNMs (i.e. TPs and REL CLs) are assumed to be syntactically more complex, the main reason being the fact that they contain a verbal element and therefore may take a subcategorisation pattern (Berlage 2014: 41-42; see also Frazier 1985; Smith 1988; Ferreira 1991; Wasow 1997; Yaruss 1999 or Gibson 2000). In line with the literature, being more complex should imply taking a higher rate of plural verb number.
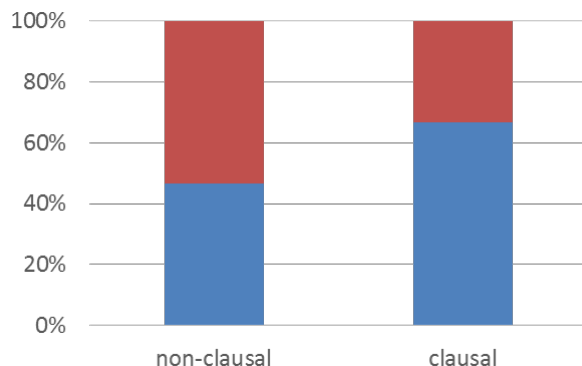
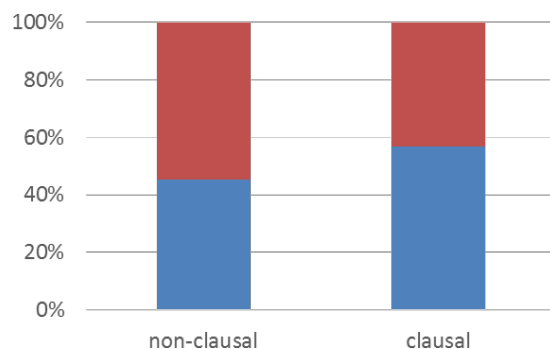**Figure 6**. Verbal agreement with each type of PNM in American English

Figure 7. Verbal agreement with each type of PNM in American English

The analysis of their corresponding patterns of agreement, however, confirms that complexity does not correlate with a higher rate of plural agreement in the constructions under scrutiny since clausal PNMs take a significantly lower rate of plural agreement in both varieties ($\chi^2(1)$, p<.0001). This finding suggests that the preference for the use of the (unmarked) singular number could be a resource to ease the cognitive processing of these complex N_{COLL}-*of*-N_{PL} constructions. Neither conclusive results nor support has been found yet and thus this issue calls for further consideration.

5 Conclusions

The research reported in this paper has demonstrated that N_{COLL}-*of*-N_{PL} subjects show significant variation in verbal agreement in Present-Day British and American English as well as in other regional varieties.

As for syntactic complexity, the results have yielded further support to my prior observations and, in turn, have further refuted earlier literature on collective nouns. The data have illustrated how in these particular constructions syntactic distance and complexity correlate with progressively lower rates of plural agreement in both British and American English. Despite the significance of this finding, it remains to be seen the effects of distance on the other regional varieties in GloWbE.

References

Algeo, J. (2006). *British or American English? A Handbook of Word and Grammar Patterns*. Cambridge: Cambridge University Press.

Bauer, L. (1994). *Watching English change.* New York: Longman.

Berlage, E. (2014). *Noun Phrase Complexity in English*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

BNC = British National Corpus <http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php>

COCA = Corpus of Contemporary American English <http://corpus.byu.edu/coca>

Corbett, G. G. (1979). "The agreement hierarchy". *Journal of English Linguistics 15*, 203–224.

Dekeyser, X. (1975). *Number and Case Relations in 19th c. British English*: *A Comparative Study of Grammar and Usage*. Belgium: De Nederlandsche Boekhandel.

Depraetere, I. (2003). "On verbal concord with collective nouns in British English". *English Language and Linguistics 7*(1), 85–127.

Fernández-Pena, Y. (2015). "Verbal agreement with collective noun-based constructions: Syntactic and lexical implications of *of*-dependents". In A. Lázaro Lafuente, M. D. Porto Requejo (eds.), *English and American Studies in Spain: New Developments and Trends*. Alcalá de Henares: University of Alcalá, 206–217.

Fernández-Pena, Y. (2017). "Patterns of verbal agreement with collective nouns taking plural *of*-dependents: A corpus-based analysis on syntactic distance". To be published in *Corpora 12*(2).

Ferreira, F. (1991). "Effects of length and syntactic complexity on initiation times for prepared sentences". *Journal of memory and Language 30*(2): 210–233.

Frazier, L. (1985) "Syntactic complexity". In D. R. Dowty, L. Karttunen, A. M. Zwicky (eds.), *Natural Language Parsing*. Cambridge: Cambridge University Press, 129–189.

Gibson, E. (2000). "The Dependency Locality Theory: A distance-based theory of linguistic complexity". In A. Marantz, Y. Miyashita, W. O'Neil (eds.), *Image, Language, Brain. Papers from the First Mind Articulation Symposium*. Cambridge, Mass.: MIT, 95–126.

GloWbE = Corpus of Global Web-based English (Dependency Bank 2.0; 2010-2013 H. M. Lehmann and G. Schneider; University of Zurich) <http://www.es.uzh.ch/services/corpling.html>

Huddleston, R., Pullum, G. K. et al. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hundt, M. (2006). "The committee has/have decided…: On concord patterns with collective nouns in inner- and outer-varieties of English". *Journal of English Linguistics 34*(3): 206–232.

Hundt, M. (2009). "Concord with collective nouns in Australian and New Zealand English". In P. Peters, P. Collins, A. Smith (eds.), *Comparative Studies in Australian and New Zealand English: Grammar and beyond*. Amsterdam: John Benjamins, 207–224.

Jantos, S. (2010). "Agreement in educated Jamaican English: A corpus-based study of spoken usage in ICE-Jamaica". In H. Dorgeloh, A. Wanner (eds.), *Syntactic Variation and Genre*. Berlin: Mouton de Gruyter, 305–331.

Kachru, B. B. (1985). "Standards, codification, and sociolinguistic realism: The English language in the outer circle". In R. Quirk, H. Widdowson (eds.), *English in the world: Teaching and learning of Language and Literature*. Cambridge: Cambridge University Press, 11–30.

Levin, M. (2001). *Agreement with Collective Nouns in English*. Lund: Lund Studies in English.

Smith, C. S. (1988). "Factors of linguistic complexity and performance". In A. Davison, G. M. Green (eds.), *Linguistic Complexity and Text Comprehension: Readability Issues Reconsidered*. Hillsdale, NJ: Lawrence Erlbaum, 247–79.

Trudgill, P., Hannah, J. (2008). *International English. A Guide to Varieties of Standard English*. London: Hodder Education.

Wasow, T. (1997). "Remarks on grammatical weight". *Language Variation and Change* 9 (1): 81–105.

Yaruss, J. S. (1999). "Utterance length, syntactic complexity, and childhood stuttering". *Journal of Speech, Language, and Hearing Research 42*(2): 329–344.