"I REGRET LYING" VS. "I REGRET THAT I LIED": VARIATION IN THE CLAUSAL COMPLEMENTATION PROFILE OF REGRET IN AMERICAN AND BRITISH ENGLISH

"I REGRET LYING" VS. "I REGRET THAT I LIED": VARIACIÓN EN EL PERFIL DE COMPLEMENTACIÓN CLAUSAL DE REGRET EN INGLÉS AMERICANO Y BRITÁNICO

RAQUEL P. ROMASANTA
Universidad de Vigo
raquel.romasanta@uvigo.es

Abstract

The historical development and change of the English complementation system has received a great deal of attention in recent years, but work remains to be done on Present-day English. Previous studies on the complement-taking predicate REGRET have shown that in British English the choice between a *that*-clause and the gerundial *-ing* is non-categorical or probabilistic, with the speaker being able to choose between them. This non-categorical variation is the focus of the present article, which aims to identify any existing differences in the clausal complementation profile of REGRET in British and American English, as well as any linguistic variables that might determine speaker choice.

Keywords: complementation, Present-day English, cognitive complexity, British English, American English.

Resumen

El desarrollo y el cambio histórico en el sistema de complementación del inglés ha recibido mucha atención, pero queda trabajo por hacer en relación con el estudio del inglés actual. Estudios previos del verbo REGRET han demostrado que en inglés británico la elección entre una oración de complemento con *that* y una oración de

gerundio en *-ing* es no categórica o probabilística y el hablante puede elegir entre ellas. Esta variación no categórica es el tema central del presente artículo, que tiene como objetivo identificar las diferencias existentes en el perfil de complementación clausal del verbo REGRET en el inglés británico y americano, así como las variables lingüísticas que pueden determinar la elección del hablante.

Palabras clave: complementación, inglés actual, complejidad cognitiva, inglés británico, inglés americano.

1. Introduction

Clausal complementation is a popular and productive area of research and it has been explored in depth in many diachronic studies in the field of English historical linguistics (see Rohdenburg 1995, 1996, 2006; Fanego 1996a, 1996b, 1996c, 1998, 2007, 2016, among others). These studies focus especially on a series of changes undergone by the English clausal complementation system over previous centuries, frequently referred to as the Great Complement Shift. Some of the major changes in this shift are the evolution of the gerund from being nominal in the early periods, to acquiring full verbal properties by Late Modern English (Fanego 2007), and its establishment "as a second type of non-finite complement" (Rohdenburg 2006: 143) at the expense of infinitives and *that*-clauses. However, despite extensive research in the field, there is room for further studies of complementation in Present-day English, "where comparatively little work has been done" (Fanego 2007: 161).

This article is an attempt to contribute to existing research by looking at the complementation profile of REGRET, a verb that has been shown to exhibit variation between finite *that*-clauses and non-finite *-ing* clauses in earlier periods of English (Heyvaert and Cuyckens 2010; Cuyckens et al. 2014; Romasanta 2017, 2019, 2021a, 2021b). In particular, I aim to explore differences in the alternation between *that*- and *-ing* complement clauses with the verb REGRET in American English (AmE) and British English (BrE), as in the examples shown in (1) below, as well as the factors that may influence a speaker's choice of one pattern over the other, as attested in data from the *Corpus of Global Web-Based English* (GloWbE; Davies 2013).

(1)   a. *Will I <u>regret</u> leav**ing** my current job with its strong sense of purpose and stability, even if it doesn't pay as well and even if I have advanced as far as I can there?* (GloWbE-US G)

        b. *I <u>regret</u> **that** I can't offer chocolate chip ones online.* (GloWbE-GB B)

The remainder of the article is organized as follows. In section 2, I present a brief account of the previous literature on variation and change in clausal complementation, and set out the research questions to be addressed. In section 3, I describe the methodology adopted and, in section 4, I present and discuss the findings. Finally, I summarize the main conclusions in section 5.

2. Background: Variation in Clause Complementation

The complementation system in English has undergone very extensive restructuring over previous centuries, commonly known as the Great Complement Shift (see Rohdenburg 1995, 1996, 2006, 2015; Fanego 1996a, 1996b, 1996c, 2004a, 2004b, 2007, 2010, 2016; Rudanko 2011; De Smet 2013). Fanego (2016: 85) discusses both major and minor restructuring changes, including the spread of infinitive clauses at the expense of finite clauses (see Rohdenburg 1995); the rise and development of *for NP to*-infinitives, as in *it was easy for him to go* (Fischer 1988; De Smet 2013: 73-101); the progressive development of verbal features of gerundives since Early Middle English times (e.g., *weakening the body by too much abstinence*) and their establishment as a second type of non-finite complement alongside infinitives (Fanego 1996b, 2004a).

One of the verbs that illustrates these restructuring over previous centuries and the one of interest to us here is REGRET and its possible complementation patterns. FrameNet, a lexical database of English which annotates examples according to their use in actual texts and provides information on their meaning, usage and valence, describes five patterns: noun phrase (NP), as in example (2), *wh*-clauses, example (3), *that*-clauses, with or without a complementizer, examples (4) and (5), *-ing*-clauses, with or without an expressed subject, examples (6) and (7), and *to*-infinitive clauses, example (8).

(2) *He regrets the fashionable attacks on critics.* (FrameNet)

(3) *"Do you regret what you've done?" she asked, moving around so that she was able to look at his face.* (FrameNet)

(4) *We regret that tickets can not be exchanged.* (FrameNet)

(5) *She regretted she would not be seeing Urquhart again for only one reason, or so she told herself.* (FrameNet)

(6) *I deeply regret my being unable to be with you and to join the memorial service and the dedication together with the related activities.* (FrameNet)

(7) *He regretted leaving Myeloski alone in the afternoon.* (FrameNet)

(8) *Now I regret to say that the only political party that actually entered into negotiations on this occasion was the Conservative Party.* (FrameNet)

39

As can be seen in examples (7) and (8), the verb REGRET belongs to the set of catenative verbs known as retrospective verbs, as is also the case with REMEMBER and FORGET. This means that the alternation between *to*-infinitive and *-ing* clauses has a functional differentiation and thus cannot be used interchangeably. The infinitive has a prospective or future meaning (*I regret to tell you that John stole it*, which can also be expressed as *I regret that I am about to tell you that John stole it*; Quirk et al. 1985: 1193), and the *-ing* construction has a retrospective or anterior meaning, as in *I regret lying*, which can be paraphrased as *I regret that I lied* (Quirk et al. 1985: 1193; see also Fanego 1996a, 1996b, 1996c; Mair 2006).

Whereas non-finite patterns seem to have a semantically dependent distribution, the alternation between finite *that*-clauses and non-finite *-ing* clauses is non-categorical or probabilistic (Heyvaert and Cuyckens 2010; Cuyckens et al. 2014; Romasanta 2017, 2019, 2021a, 2021b). That is, speakers' choices between the two structures seem to be motivated by factors other than meaning. This non-categorical or probabilistic variation is the focus of the present article, specifically in contemporary American and British English. Cuyckens et al. (2014) consider the non-categorical differentiation between *that*-clauses and non-finite clauses with the verbs REGRET, REMEMBER, and DENY in Late Modern English. They examine the changes that have occurred in complement choice over time and try to determine the semantic and syntactic factors that condition the type of complement clause used. They find that, with REGRET, there was a slight increase in the use of finite patterns (*that*- and zero *that*-clauses) over non-finite patterns (*-ing* clauses and *to*-infinitives). For all three verbs, Cuyckens et al. (2014) conclude that a number of factors determining speaker choice disfavor the use of non-finite patterns: the use of state verbs in the complement clause, the presence of intervening material, a complex noun phrase as subject of the complement clause, non-coreferential subjects, inanimate subjects in the complement clause, and an anterior temporal relation between the complement clause and the main clause. Some of these syntactic factors that have an effect on complement alternation relate to the Complexity Principle (Rohdenburg 1995, 1996, 2006), which expresses a correlation between cognitive complexity and grammatical explicitness (Rohdenburg 2006: 146). Rohdenburg describes the Complexity Principle as, "in the case of more or less explicit constructional options, the more explicit one(s) will tend to be preferred in cognitively more complex environments" (1996: 151). Generalizations related to linguistic complexity include, for example, the fact that negative sentences are cognitively more complex than their affirmative counterparts, that the presence of any intervening material between the main and subordinate clauses increases the processing burden, and that the processing complexity of passives is greater than their active counterparts (Rohdenburg 1996, 2006).

In the case of complementation, finite clausal structures are more explicit than non-finite ones because they explicitly mark tense, agreement and modality, and have an explicit subject and a complementizer (that is, when it has one; see example (9) below). All this makes it more explicit and transparent and easier to learn and process than non-finite structures (Givón 1985; Ortega 2003; Steger and Schneider 2012; Di Domenico 2017; Green 2017).

(9)   a. *I regret **that** you left the navy.* (GloWbE-GB G)
      b. *I regret you leaving the navy.*

For the present study, I decided to compare variations found in the American English and British English components of the *Corpus of Global Web-Based English* (GloWbE, Davies 2013). Different terminology has been adopted to refer to these two varieties, such as, national varieties (e.g., Algeo 2006; Rohdenburg and Schlüter 2009b), mainstream varieties (Schneider 2007; Williams et al. 2015), L1 varieties (Hoffmann and Siebers 2009), supranational varieties (Mair 2013), and metropolitan varieties (Aceto 2004), among others. As Rohdenburg and Schlüter claim, "at many levels of description, British-American contrasts are widely recognized" (2009b: 1). For instance, Rohdenburg's study on nominal and prepositional complementation refers to American English as favoring "the formally less explicit or simpler option over its more complex variant" (2009: 194). In the current article, this hypothesis will be tested for clausal complementation. Another example of the contrast between the grammars of American and British English is non-finite sentential complementation (Vosberg 2009). Vosberg shows that even though both varieties "follow the same trends in the development of non-finite complement variants [they do so] at clearly different speeds" (2009: 226). In general, American English is considered more innovative than British English (Leech et al. 2009; Mair 2013), and I seek to explore whether differences in the use of REGRET between these two varieties provides evidence of this. In particular, I seek to answer the following research questions:

- What are the differences and similarities between the clausal complementation profiles of the verb REGRET in American and British English?

- What are the differences and similarities within the two varieties between the General and the Blogs corpus subdivisions?

- Which linguistic factors are relevant for speaker choice among the complementation patterns available?

- Does the Complexity Principle (Rohdenburg 1996) affect the complement choice to any extent?

3. Data and Methodology

3.1. Data Selection

As noted above, the data for this study have been retrieved from the GloWbE corpus, using its website interface. This is an online corpus released in 2015 with data taken from the Internet from the years 2012-2013 (Davies and Fuchs 2015; Lange and Leuckert 2020). It contains 1.9 billion words from 20 different countries in which English is spoken, either as L1 (e.g., the United States, Canada, Great Britain) or as L2/L3 (e.g., India, Sri Lanka, Philippines, Nigeria). The two components used for this study contain a total of 774,424,429 words (386,809,355 words in the American English component and 387,615,074 words in the British English component). The texts are divided into two categories: the Blogs corpus subdivision, which accounts for about 60% of the corpus, and the General subdivision, which accounts for 40% of the corpus and contains web-based materials such as newspapers, magazines, and company websites. This distinction between Blogs and General, intended to provide a contrast between more speech-based texts and more written-based and formal ones, has been questioned in previous research. Loureiro-Porto (2017), for instance, compares GloWbE to the *International Corpus of English* (ICE; Greenbaum 1996) and argues that, in terms of orality and informality, both sections of GloWbE, Blogs and General, are very similar.

The choice of the GloWbE corpus for the present study is informed by its size. For example, the ICE corpora is smaller and a search of its British and Indian components yielded only 41 and 25 examples of REGRET. The greater number of words in GloWbE allows us to study low-frequency structures, as is the case with clausal complementation (Davies 2012: 162), as well as low-frequency lexical items such as REGRET (Romasanta 2021a). Another reason for the selection of this corpus is its inclusion of numerous varieties of English from around the world, which have already been the focus of research and will continue to be so (see, for example, Romasanta 2017, 2019, 2021a, 2021b).

However, the use of GloWbE also brings with it a number of difficulties, and I have tried to minimize these as follows. First, there is some duplication of examples. In order to identify and discard these, I sorted all the examples of each variety alphabetically. Secondly, I identified some imperfections in the tagging of the corpus which yielded false positives. I dealt with these by manually cleaning any false positives that I had retrieved through the search for the verb REGRET (regret*_v*); that is, any examples in which REGRET was not a verb but was either used as a noun or an adjective (as in examples (10) and (11) below) were excluded.

> (10) *I would say this record displays a wide range of themes —family, love, <u>regret</u>, fear, youth,…* (GloWbE-GB B)

(11) …*whose style of preaching you find painfully below that of his* <u>*regretted*</u> *predecessor?*
    (GloWbE-US G)

All the instances of the verb REGRET retrieved in American English and British English were included in an Excel spreadsheet, and all false positives or duplicated examples discarded. I then took a random sample of 2,000 hits from each variety. The total number of examples for each variety is proportional in terms of text type, in that I took 1,000 from General and 1,000 from Blogs.

3.2. Data Coding

The 4,000 examples selected needed further manual cleaning and pruning. For example, I decided to exclude the examples in which REGRET takes a compound or coordinated complement clause, because it is not possible to examine the relevant linguistic factors in such clauses. This limitation is illustrated below in the discussion of the factors selected.

In order to establish the final number of valid tokens, and to limit the scope of variation under investigation, I considered two factors:

a. COMPLEMENTATION TYPE. This is the dependent variable and has two values:
   i. *that*
   ii. (S) *-ing*

b. TEMPORAL RELATION. This factor refers to the temporal relation between the time expressed by the verb in the complement clause and the time expressed by the verb in the main clause.
   i. anterior (example (12) below)
   ii. simultaneous, as in (13).

43

(12) a. *The United States government* <u>*regrets*</u> **that** *some individuals have abused their right to free speech by showing disrespect for other nations,…* (GloWbE-US G)

b. *I* <u>*regret*</u> *tak**ing** part in the fighting now because of what has happened to me.* (GloWbE-GB G)

(13) a. *I* <u>*regretted*</u> **that** *I hadn't any cash with me.* (GloWbE-GB B)

b. *I* <u>*regret*</u> *leav**ing** my friends here.* (GloWbE-US G)

The total number of valid tokens, that is, examples of the verb REGRET followed by a *that-* or *-ing* clause with anterior or simultaneous temporal relation, is shown in Table 1 below. These examples were then coded for a number of variables, in order to shed light on the factors that determine complement choice.

	General	Blogs	TOTAL
AmE	247	287	534
BrE	289	289	578

Table 1. Total number of viable examples of REGRET in American and British English

Among the factors to be taken into consideration, I included some external ones related to the text type and the variety of English:

c. TEXT TYPE. General (G) and Blogs (B).

d. VARIETY OF ENGLISH. American English (AmE) and British English (BrE).

As for language-internal factors, Table 2 below provides a summary of the intra-linguistic factors included in the analysis.

Main clause	Complement clause	Relationship between clauses
e. MC TYPE OF SUBJECT	f. CC VOICE g. CC VERBAL MEANING h. CC TYPE OF SUBJECT i. CC SUBJECT ANIMACY j. CC COMPLEXITY k. CC NEGATION	l. SUBJECT COREFERENTIALITY m. INTERVENING MATERIAL

Table 2. Summary of the intra-linguistic factors

The internal linguistic variable (e) below concerns the main clause (MC):

e. MC TYPE OF SUBJECT. This factor is important in terms of examining the influence of Rohdenburg's Complexity Principle (1996) on the choice of complementation patterns.

    i. pronoun: 1st, 2nd, 3rd person personal pronouns, and *wh*-pronoun (example (14) below);

    ii. simple noun phrase: one or two words, example (15);

    iii. complex noun phrase: more than two words, as in (16);

    iv. none: an explicit subject in the main clause is missing, see example (17).

(14) ***We regret** greatly that the memorandum did not measure up to what the Director...* (GloWbE-US G)

(15) ***Democrats** will regret embracing the expansion of executive power because...* (GloWbE-US B)

(16) *a new study has revealed that **nearly a third of people** <u>regret</u> getting them done.* (GloWbE-GB B)

(17) *Show facts: do not <u>regret</u> missing this show, because the music was…* (GloWbE-US G)

Regarding the complement clause (CC), I considered the following factors:

f. CC VOICE. This factor has been considered to play a role in the selection of competing options within the study of Rohdenburg's Complexity Principle (1996). Two values are distinguished:
    i. active, as in (18);
    ii. passive, see example (19);
    iii. copular/intransitive, as in (20).

(18) *Nonetheless, he doesn't <u>regret</u> try**ing** to make the switch.* (GloWbE-US G)

(19) *It is so common to hear how people <u>regret</u> not hav**ing** their wedding filmed professionally after the event…. when it's too late* (GloWbE-GB B)

(20) *Some days I <u>regret</u> **that** I did not go into another line of work.* (GloWbE-US G)

g. CC VERBAL MEANING. The verb of the complement clause may denote a state (example (21)), or an event/action, as in (22), following Quirk et al.'s (1985: 201) classification.

(21) *I <u>regret</u> **being** alive still.* (GloWbE-US G)

(22) *All in all, I <u>regret</u> **paying** for this game.* (GloWbE-GB B)

This is one of the factors which cannot be studied in compound or coordinated complement clauses because the verbs in the different complement clauses may have different meanings. This is also the case for the type of subject of the complement clause, subject animacy, complexity, negation, and coreferentiality (sections h, i, j, k, l below). Therefore, as already noted, compound or coordinated complement clauses are excluded from the analysis of the factors.

h. CC TYPE OF SUBJECT. As well as in the main clause, this factor is important in order to consider the influence that the Complexity Principle may have on the choice of complement clause type (Rohdenburg 1996). It has two levels:
    i. complex noun phrase (CNP), as in example (23). A complex noun phrase as subject of the complement clause contains more than one word.
    ii. Other, as in example (24).

(23) *I' too, <u>regret</u> that **a misleading and rather vexatious attempt to divert attention from the potential importance of the overall result** was made.* (GloWbE-US B)

(24) *As an atheist, I <u>regret</u> that **I** cannot consign that malignant miserable…* (GloWbE-US G)

45

i. CC SUBJECT ANIMACY. The subject in the complement clause may be animate, as in (24), or inanimate, as in (23) above. However, there are examples of collective nouns which can be considered either animate or inanimate. Sometimes plural agreement with the verb stresses the personal individuality within the group, as in (25), and examples are therefore classified as animate. In some other cases, singular agreement with the verb stresses the non-personal collectivity of the group, being understood as a company, association, team, etc., and therefore examples are classified as inanimate, as in (26) (Quirk et al. 1985: 316-317). There are, however, some problematic examples in which the verb in the complement clause semantically requires an animate subject (see example 27), with a volitional ('be willing to') meaning. In such cases, the subject is considered to be animate, regardless of the agreement distinction.

> (25) *I <u>regret</u> that **arsenal** are so lacking in these qualities at present.* (GloWbE-GB B)

> (26) *I <u>regret</u> that **the Court** has stopped short of this holding indispensable to…* (GloWbE-US G)

> (27) *I <u>regret</u> very much that the **United Kingdom** was not willing to join…* (GloWbE-GB G)

j. CC COMPLEXITY. This factor is important in order to examine the potential influence of the Complexity Principle (Rohdenburg 1996). This variable was operationalized as the number of graphemic words. In example (25) above, for instance, the number of words would be 10.

k. CC NEGATION. Rohdenburg (2006, 2015) considers the presence of a negative marker as adding complexity to the sentence and, therefore, as potentially conditioning the speaker's choice of one complementation pattern over the others. Romasanta (2021c) studies the effect of different negative markers on the alternation between *that-* and *-ing* clauses and *that-* and zero-complement clauses across 16 varieties of English and finds that, when considering *that-* and *-ing* complement clauses, (i) both *not*-negation and *no*-negation trigger the use of finite complement clauses, and (ii) this preference for *that*-clauses is stronger with *no*-negation. Here I look at the effect of the presence of any type of negative marker (*not*, *n't*, *never*, *neither*, *nobody*, *no*, *none*, *nor*, *nothing*, *nowhere*) on the complement clause. This factor has two levels:

> i. affirmative, example (28);

> ii. negative, as in (29).

> (28) *Does Maher <u>regret</u> spending 14 years in the closet professionally?* (GloWbE-US G)

> (29) *I had always encouraged him to at least try before he got too old and <u>regretted</u> **never** even making an attempt at it.* (GloWbE-US G)

In addition, I considered some factors related to the relationship between the main clause and the complement clause, namely:

l. SUBJECT COREFERENTIALITY between the subject of the main clause and the subject of the complement clause. The two values are coreferential (see example (30) below) and non-coreferential, as in (31).

   (30) *Given the sensitive nature of this matter* **we** *regret that* **we** *are unable to provide further details at this stage.* (GloWbE-GB B)

   (31) **They** *would be* *regretting* **the press** *being there, not the usual come on to them.* (GloWbE-GB G)

m. INTERVENING MATERIAL. Rohdenburg (1995: 376) considers the presence of intervening material as an important factor that adds complexity to the structure and that may therefore influence the speaker's choice of complement clause. It indicates the number of words that occur between the verb REGRET and the first word of the complement clause, as in example (32).

   (32) *You will* *regret* **more and more each day** *this non-suit is in OUR White House* (GloWbE-US B)

4. Results

Figure 1 shows the distributions of finite *that*-clauses and non-finite *-ing* clauses in the British and American Englishes samples. We see that these distributions are very similar, with a clear preference for *-ing* clauses, 73.01% of *-ing* clauses in British English and 73.03% in American English. This similarity in distribution may indicate that the two varieties are not so different in terms of their complementation systems, at least for the verb REGRET.
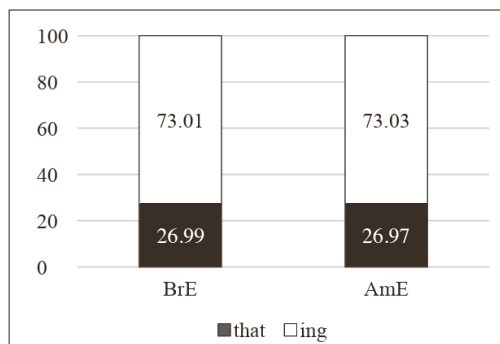

Figure 1. Distribution of finite *that*-clauses and non-finite *-ing* clauses complementing REGRET in British and American English

Table 3 summarizes the extra- and intra-linguistic factors analyzed and their distribution in the data. Figure 2 presents the distribution of the individual factors considering the complementation type, *that* or *-ing* clause, and the relative effect that the factors have on the choice between the two patterns. I use Chi-square and Cramer's V tests to test for independence between dependent and independent factors and describe the type of relationship, if any (Field et al. 2012: 818).

Factor	Distribution
a. COMPLEMENTATION TYPE	N = 300 **that** (26.98%); N= 812 *-ing* (73.02%)
b. TEMPORAL RELATION	N = 966 anterior (86.87%); N = 146 simultaneous (13.13%)
c. TEXT TYPE	N = 536 General (48.2%); N = 576 Blogs (51.8%)
d. VARIETY OF ENGLISH	N = 578 BrE (51.98%); N = 534 AmE (48.02%)
e. MC TYPE SUBJECT	N = 91 none (8.18%); N = 176 NP (15.83%); N = 561 pron1 (50.45%); N = 109 pron2 (9.8%); N = 175 pron3 (15.74%)
f. CC VOICE	N = 781 active (70.23%); N = 50 passive (4.5%); N = 281 copular (25.27%)
g. CC VERBAL MEANING	N = 946 event (85.07%); N = 166 state (14.93%)
h. CC TYPE OF SUBJECT	N = 109 CNP (9.8%); N = 1,003 other (90.2%)
i. CC SUBJECT ANIMACY	N = 984 animate (88.49%); N = 128 inanimate (11.51%)
j. CC COMPLEXITY	*Mean* = 8.79; *Min* = 1; *Max* = 87
k. CC NEGATION	N = 741 affirmative (66.64%); N = 371 negative (33.36%)
l. SUBJECT COREFERENTIALITY	N = 934 coreferential (83.99%); N = 178 non-coreferential (16.01%)
m. INTERVENING MATERIAL	*Mean* = 0.11; *Min* = 0; *Max* = 9

Table 3. Descriptive statistics of the extra- and intra-linguistic factors described in Section 3.2

The following values show a significantly stronger preference for *that*-clauses: simultaneous temporal relation between the main clause and the complement clause, state verbs, complex noun phrases and inanimate entities as subjects in the complement clause, passive and negative complement clauses, non-coreferential subjects between the clauses, and longer complement clauses. There are also significant differences in the distribution depending on the type of subject in the main clause, with a stronger preference for *that*-clauses with first person personal pronouns, noun phrases, and no subject. The factors TEXT TYPE, VARIETY, and INTERVENING MATERIAL do not seem to have a significant effect on the choice of complementation.

In the results for the Cramer's V test for the measurement of the effect size, absolute values of 0 mean no association and absolute values of 1 mean perfect association.

Text type

(a) $x^2 (1) = 2.17$, p > .001

Cramer's V = 0.04

Variety

(b) $x^2 (1) = 3.99e^{-30}$, p > .001

Cramer's V = $5.99e^{-17}$

MC type subject

(c) $x^2 (4) = 56.88$, p < .001

Cramer's V = 0.22

CC voice

(d) $x^2 (2) = 103.97$, p < .001

Cramer's V = 0.31

CC verbal meaning

(e) $x^2 (1) = 78.45$, p < .001

Cramer's V = 0.27

CC type subject

(f) $x^2 (1) = 275.85$, p < .001

Cramer's V = 0.5

CC subject animacy

(g) $x^2 (1) = 279.47$, p < .001

Cramer's V = 0.5

CC negation

(h) $x^2 (1) = 30.29$, p < .001

Cramer's V = 0.17

Subject coreferentiality

(i) $x^2 (1) = 484.67$, p < .001

Cramer's V = 0.66

Temporal relation

(j) $x^2 (1) = 149.47$, p < .001

Cramer's V = 0.37

CC complexity

(k) Binary logistic regression estimate = -0.14, p < .001

Intervening material

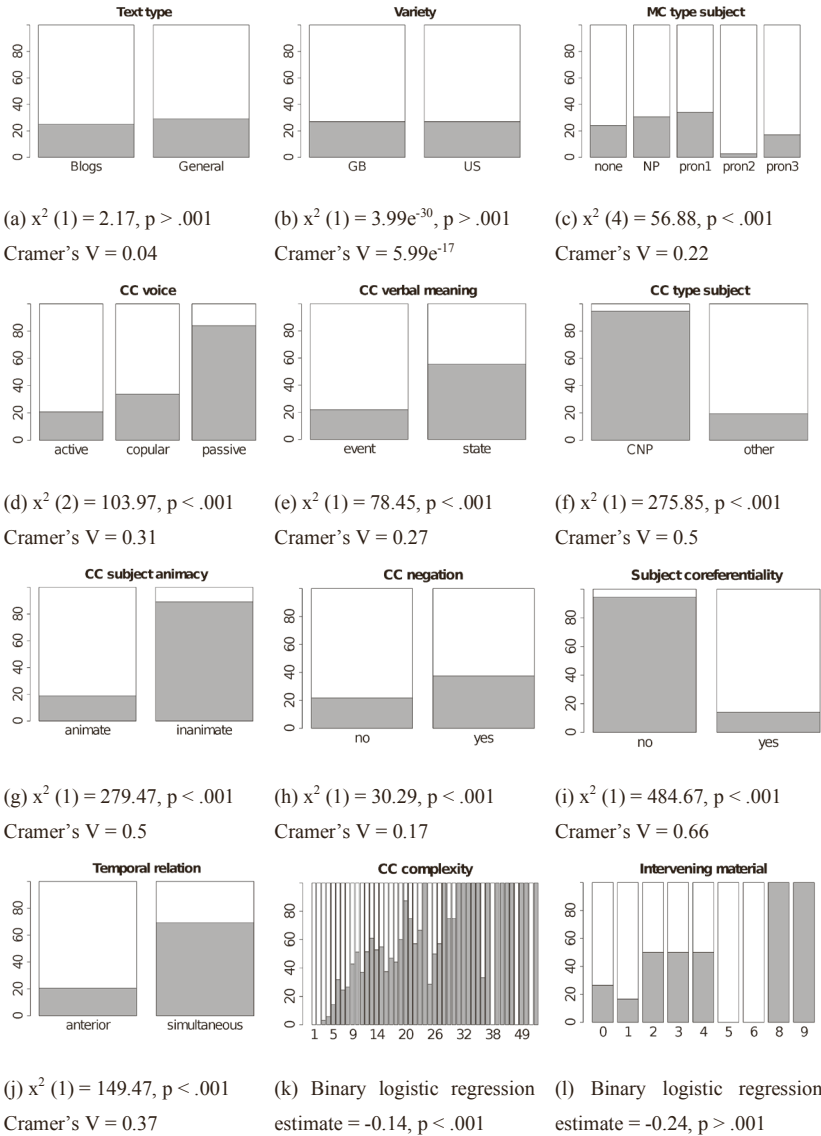(l) Binary logistic regression estimate = -0.24, p > .001

Figure 2. Distribution of extra- and intra-linguistic factors according to complementation type

49

Here we see that the factors CC VOICE, CC SUBJECT TYPE, CC SUBJECT ANIMACY, and SUBJECT COREFERENTIALITY have a strong effect size, and MC TYPE SUBJECT, CC VERBAL MEANING, and TEMPORAL RELATION have a moderate effect size. All the other factors analyzed have a small effect size (TEXT TYPE, VARIETY, and CC NEGATION).

Table 4 sets out the results from the binary logistics regression model by means of the coefficients, standard errors, p-values with significance rating, odds ratio, and confidence intervals. Predicted estimates are for finite *that*-clauses. The factor with the strongest significant effect on the alternation between *that*- and *-ing* clauses with the verb REGRET, that is, the one with the largest number under the coefficients column, is the coreferentiality between the subjects of the main and complement clause (SUBJECT COREFERENTIALITY). With coreferential subjects, the preference for finite *that*-clauses decreases considerably (negative number in the coefficients column). The other factors that significantly determine this alternation are the type of subject in the main clause (MC TYPE SUBJECT), TEMPORAL RELATION, complexity

	coef.	std. error	p-value		OR	CI 95%
Intercept	-1.69	0.80	0.03	*	0.18	(-3.25, -0.09)
TEXT_TYPE: General	0.29	0.22	0.18		1.34	(-0.12, 0.71)
VARIETY: AmE	0.24	0.22	0.27		1.27	(-0.18, 0.67)
MC_TYPE SUBJECT (default: none)						
MC TYPE SUBJECT: NP	-0.92	0.44	0.04	*	0.40	(-1.79, -0.05)
MC TYPE SUBJECT: pron1	-0.33	0.37	0.37		0.72	(-1.03, 0.41)
MC TYPE SUBJECT: pron2	-2.56	0.89	0.004	**	0.08	(-4.62, -1.03)
MC TYPE SUBJECT: pron3	-0.73	0.45	0.10		0.48	(-1.61, 0.16)
CC_VOICE (default: active)						
CC_VOICE: copular	-0.07	0.26	0.80		0.94	(-0.59, 0.43)
CC_VOICE: passive	1.00	0.74	0.18		2.73	(-0.42, 2.47)
CC_VERBAL_MEANING: state	0.37	0.29	0.21		1.44	(-0.21, 0.93)
CC_TYPE_SUBJECT: other	0.78	0.74	0.29		2.18	(-0.65, 2.31)
CC_ANIMACY: inanimate	1.07	0.57	0.06		2.93	(-0.07, 2.18)
CC_NEGATION: negative	0.96	0.22	$1.49e^{-05}$	***	2.60	(0.53, 1.39)
SUBJECT COREFERENTIALITY: coreferential	-4.18	0.61	$9.10e^{-12}$	***	0.02	(-5.53, -3.08)
TEMPORAL_RELATION: simultaneous	1.82	0.31	$2.36e^{-09}$	***	6.19	(1.23, 2.43)
log_cc_words	1.37	0.17	$1.01e^{-15}$	***	3.94	(1.04, 1.72)
log_intervening_material	0.006	0.27	0.98		1.01	(-0.54, 0.51)
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Table 4. Summary of the binary logistic regression model together with Odds Ratio (OR) and 95% Confidence Intervals (CI) for the estimated OR

of the complement clause (log_cc_words), and CC NEGATION. From the signs in the coefficients column, we see that with second person personal pronouns and noun phrases as subjects of the main clause the tendency for the use of *that*-clauses also decreases (negative number). In contrast, a simultaneous temporal relation between the main and complement clause seems to favor the use of *that*-clauses to complement REGRET (positive numbers). The same applies to longer and negative complement clauses.
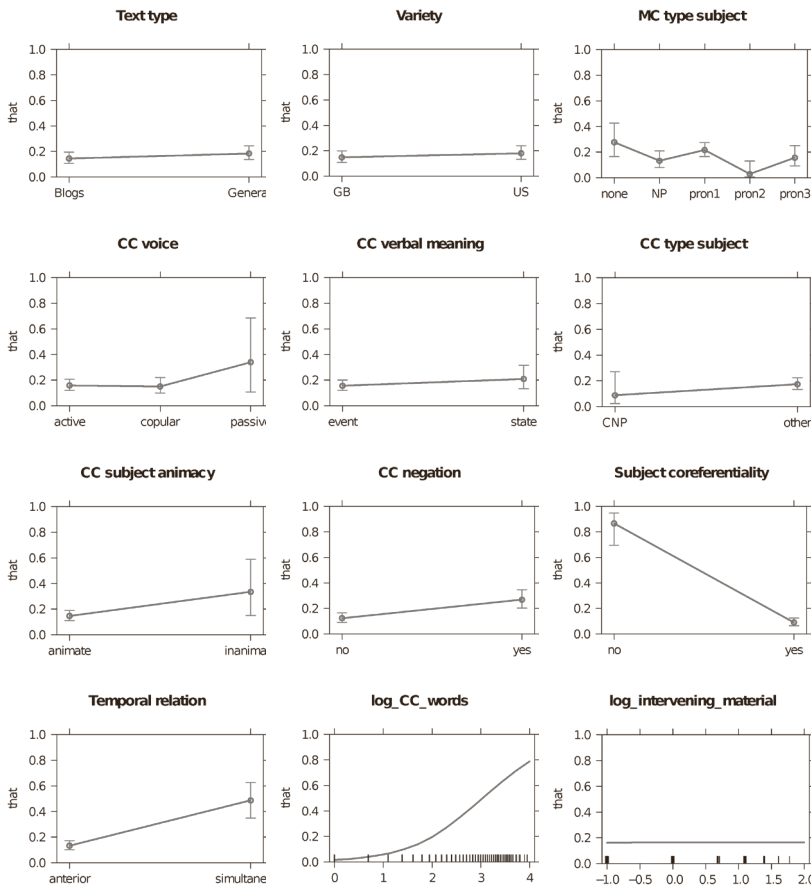
Figure 3. Effect size of the extra- and intra-linguistic factors described in Section 3.2

None of the other factors, namely TEXT TYPE, VARIETY, CC VOICE, CC VERBAL MEANING, CC TYPE SUBJECT, CC ANIMACY, and INTERVENING MATERIAL, have a significant effect on the choice. It therefore appears that the distinctions between General and Blogs and between British and American English are not significant for the present alternation between finite *that*-clauses and non-finite -*ing* clauses. The fact that the two sections of the corpus, General and Blogs, do not show significant differences in the complementation of the verb REGRET seems to provide further confirmation of the findings in Loureiro-Porto (2017) regarding the orality and informality of Blogs. In her study, the author finds that Blogs "cannot be considered equivalent to spoken material" (2017: 46) and that the use of linguistic features characteristic of oral and informal language does not differ significantly in the data.

Figure 3 graphically illustrates the effect of each factor in the model. In this figure, the 95% error bars in each factor express the uncertainty of the estimate and are marked with vertical lines for the categorical factors and with a shadow for the continuous ones. When these error bars are fairly wide, we cannot say whether there is a definite effect on the alternation. Looking at the figure, we see again that the strongest predictor seems to be SUBJECT COREFERENTIALITY, with non-coreferential subjects clearly increasing the proportion of *that*-clauses. Second person personal pronouns as subjects of the main clause prefer the use of -*ing* clauses, as do noun phrases but to a lesser extent. In contrast, a simultaneous temporal relation between the two clauses, as well as longer and negative clauses, shows a preference for *that*-clauses.

In the discussion in section 2, I considered Rohdenburg's Complexity Principle (Rohdenburg 1995, 1996, 2006). Rohdenburg (2006: 146) describes this as a correlation between cognitive complexity and grammatical explicitness, which could have an effect on different alternations in the language. Some of the results presented here are consistent with this principle in the sense that cognitive complex environments, such as negative and long complement clauses, favor the most explicit option, *that*-clauses. However, complex subjects in the main clause, such as noun phrases (NP), seem to favor the less explicit option, -*ing* clauses. Other cognitively complex environments, such as the passive voice, complex subjects in the complement clause, and the presence of intervening material between the two clauses, do not seem to have any effect on the alternation between -*ing* and *that*-clauses. Thus, we might agree with the conclusion of Cuyckens et al. that "Rohdenburg's proposed disfavouring effect cannot be generalised to all structural complexity factors, and in that sense, Rohdenburg's complexity principle does not apply as generally as commonly held" (2014: 199).

5. Conclusion

This article considers the complementation profile of the verb REGRET in two supranational varieties of contemporary English, namely American and British English, as represented in the language of the internet, in the *Corpus of Global Web-Based English* (GloWbE). It focuses on the non-categorical alternation between the finite *that*-clauses and non-finite *-ing* clauses and a number of intra- and extra-linguistic factors that could potentially have an effect on speaker choice between the two complementation patterns.

Based on previous research, the article poses four research questions. Regarding the first question on the differences and similarities between the clausal complementation profiles of the verb REGRET in American and British English, my findings show that these are not statistically significant. Both varieties show a stronger tendency for *-ing* complement clauses (73.01% of *-ing* clauses in British English and 73.03% in American English). This clear preference for *-ing* clauses seems to continue the historical spread of this pattern at the expense of finite complement clauses. Research question (2) concerns the potential differences and similarities between the two text types contained in the corpus, General and Blogs. The differences between these have not been found to be statistically significant.

Research question (3) refers to the factors that are relevant for speaker choice regarding the dichotomy between finite *that*-clauses and non-finite *-ing* patterns. The following factors turned out to have a statistical significance and seem to trigger an increased preference for *that*-clauses: simultaneous temporal relation between the main and complement clause, and longer and negative complement clauses. With second person personal pronouns and noun phrases as subjects of the main clause, and with coreferential subjects, however, the tendency for *that*-clauses decreases significantly. The effect of the following factors is not statistically significant: text type, variety of English, voice of the complement clauses, verbal meaning and type of subject in the complement clause, the animacy of the subject in the complement clause, and the presence of intervening material between the two clauses.

Regarding the potential effect of the cognitive linguistic principle that may influence the speaker's choices between the available complementation patterns considered, research question (4), the Complexity Principle (Rohdenburg 1995, 1996, 2006) postulates that an increase in the processing burden favors the use of the more explicit forms. I examined several features that are said to increase structural complexity and found some of them to have an influence on the choice of complementation pattern. When some features occur in a clause there is a

statistically significant higher proportion of *that*-clauses, the more explicit grammatical form, over -*ing* clauses. These features are the presence of negative markers in the complement clause and long complement clauses. Interestingly, however, complex subjects in the main clause, such as noun phrases (NP), increase the tendency for the more complex, less explicit -*ing* clauses. From this, we cannot draw any firm conclusions concerning the effect of the Complexity Principle in general or its effect on the clausal complementation system of the verb REGRET in particular.

To conclude, the findings of this study contribute to several areas of research. First, they confirm the historical tendency in English to replace finite patterns with non-finite -*ing* complement clauses, a tendency attested and described in numerous diachronic studies (Rohdenburg 1995, 1996, 2006, 2015; Fanego 1996a, 1996b, 1996c, 2004a, 2004b, 2007, 2010, 2016; Rudanko 2011; De Smet 2013). Secondly, the findings contribute to the existing literature on non-categorical variation in clausal complementation and the intra-linguistic factors determining the speaker's choice between multiple possibilities (Heyvaert and Cuyckens 2010; Bernaisch 2013; Cuyckens et al. 2014; Deshors 2015; Deshors and Gries 2016; Romasanta 2017, 2019, 2021a, 2021b, among others). Third, from a methodological perspective the study shows that a combination of qualitative work (the detailed manual pruning and selection of valid examples) and quantitative analysis (binary logistic regression modeling) is necessary to obtain reliable results. Fourth, it provides further evidence in support of the view that the two sections into which GloWbE is divided, General and Blogs, are not in fact very different. At the same time, it shows that the corpus is a valuable tool for the analysis of low frequency lexical items and constructions, given its large size, as long as the researcher is willing to conduct a fine-grained weeding out of repeated and mistagged examples. Finally, this article shows that variation in complementation remains a prolific area of research, especially regarding (i) the language-internal and external predictors that determine it, and (ii) the intervarietal differences that may, or may not, characterize it. As claimed by Fanego (2007), much research is needed in this area, and this claim should be extended to cover variation in other varieties of English around the world. As Schneider points out, "in the process of structural nativization, verbs begin to allow and later prefer new structures to complement them and build a complete sentence" (2007: 86). This includes outer-circle languages, already addressed in García-Castro (2018, 2019, 2020) and Romasanta (2017, 2019, 2021a, 2021b), expanding-circle varieties, and learner varieties, where cognitive factors derived from language contact in multilingual settings and second language acquisition are likely to pay a role in clausal complement choice.

Acknowledgments

Works cited

ACETO, Michael. 2004. "Eastern Caribbean English-derived Language Varieties: Phonology". In Schneider, Edgar W. (ed.) *Varieties of English: The Americas and the Caribbean*. Berlin/New York: Mouton de Gruyter: 290-311. DOI: 10.1017/CBO9780511607240

ALGEO, John. 2006. *British or American English? A Handbook of Word and Grammar Patterns*. Cambridge: Cambridge U.P.

BERNAISCH, Tobias. 2013. "The Verb-complementational Profile of Offer in Sri Lankan English". In Huber, Magnus and Joybrato Mukherjee (eds.) *Corpus Linguistics and Variation in English: Focus on Non-native Englishes*. Helsinki: Varieng. <https://varieng.helsinki.fi/series/volumes/13/bernaisch/>. Accessed February 10, 2021.

CUYCKENS, Hubert, Frauke D'HOEDT and Benedikt SZMRECSANYI. 2014. "Variability in Verb Complementation in Late Modern English: Finite vs. Non-finite Patterns". In Hundt, Marianne (ed.) *Late Modern English Syntax*. Cambridge: Cambridge U.P.: 182-204. DOI: 10.1017/CBO9781139507226.014

DAVIES, Mark. 2012. "Examining Recent Changes in English: Some Methodological Issues". In Nevalainen, Terttu and Elizabeth C. Traugott (eds.) *The Oxford Handbook of the History of English*. Oxford: Oxford U.P.: 263-287.

DAVIES, Mark. 2013. "Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries". <http://corpus.byu.edu/glowbe/>. Accessed February 8, 2017.

DAVIES, Mark and Robert FUCHS. 2015. "Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-based English Corpus (GloWbE)". *English World-Wide* 36 (1): 1-28. DOI: 10.1075/eww.36.1.01dav

DE SMET, Hendrik. 2013. *Spreading Patterns: Diffusional Change in the English System of Complementation*. New York: Oxford U.P. DOI: 10.1093/acprof:oso/9780199812752.001.0001

DESHORS, Sandra C. 2015. "A Constructionist Approach to Gerundial and Infinitival Verb Complementation Patterns in Native and Hong Kong English Varieties". *English Text Construction* 8 (2): 207-235. DOI: 10.1075/etc.8.2.04des

DESHORS, Sandra C. and Stefan Th. GRIES. 2016. "Profiling Verb Complementation Constructions across New Englishes: A Two-step Random Forest Analysis of *ing* vs. *to* Complements". *International Journal of Corpus Linguistics* 21 (2): 192-218. DOI: 10.1075/ijcl.21.2.03des

DI DOMENICO, Elisa. 2017. "Introduction". In Di Domenico, Elisa (ed.) *Syntactic Complexity from a Language Acquisition Perspective*. Cambridge: Cambridge Scholars Publishing: 1-22.

FANEGO, Teresa. 1996a. "On the Historical Development of the English Retrospective Verbs". *Neuphilologische Mitteilungen* 97: 71-79.

FANEGO, Teresa. 1996b. "The Development of Gerunds as Objects of Subject-control Verbs in English (1400-1760)". *Diacronica* 13: 29-62. DOI: 10.1075/dia.13.1.03fan

FANEGO, Teresa. 1996c. "The Gerund in Early Modern English: Evidence from the Helsinki Corpus". *Folia Linguistica Historica* 17 (1-2): 97-152. DOI: 10.1515/flih.1996.17.1-2.97

FANEGO, Teresa. 1998. "Developments in Argument Linking in Early Modern English Gerund Phrases". *English Language and Linguistics* 2: 87-119. DOI: 10.1017/S1360674300000708

FANEGO, Teresa. 2004a. "On Reanalysis and Actualization in Syntactic Change: The Rise and Development of English Verbal Gerunds". *Diachronica* 21: 5-55. DOI: 10.1075/dia.21.1.03fan

FANEGO, Teresa. 2004b. "Some Strategies for Coding Sentential Subjects in English: From Expectation to Grammaticalization". *Studies in Language* 28: 321-361. DOI: 10.1075/sl.28.2.03fan

FANEGO, Teresa. 2007. "Drift and Development of Sentential Complements in British and American English from 1700 to the Present Day". In Pérez-Guerra, Javier, Dolores González-Álvarez, Jorge Luis Bueno-Alonso and Esperanza Rama-Martínez (eds.) "*Of Varying Language and Opposing Creed": New Insights into Late Modern English*. Bern: Peter Lang: 161-235.

FANEGO, Teresa. 2010. "Variation in Sentential Complements in Eighteenth- and Nineteenth-Century English: A Processing-based Explanation". In Hickey, Raymond (ed.) *Eighteenth-century English*. Cambridge: Cambridge U.P.: 200-220. DOI: 10.1017/CBO9780511781643.012

FANEGO, Teresa. 2016. "The Great Complement Shift Revisited: The Constructionalization of ACC-*ing* Gerundives". *Functions of Language* 23 (1): 84-119. DOI: 10.1075/fol.23.1.05fan

FIELD, Andy, Jeremy MILES and Zoë FIELD. 2012. *Discovering Statistics Using R*. London: SAGE.

FISCHER, Olga. 1988. "The Rise of the *for NP to V* Constructions: An Explanation". In Nixon, Graham and John Honey (eds.) *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*. London: Routledge: 67-88. DOI: 10.4324/9781003074687-7

FrameNet. 2015. "FrameNet Project". <https://framenet.icsi.berkeley.edu>. Accessed May 3, 2017.

GARCÍA-CASTRO, Laura. 2018. "Variation in Verbal Complementation Patterns in Nativised Varieties of English: The Case of *Remember* in Indian English". In Ferrández San Miguel, María and Claus-Peter Neuman (eds.) *Taking Stock to Look Ahead: Celebrating Forty Years of English Studies in Spain*. Zaragoza: Prensas Universitarias de la Universidad de Zaragoza: 207-213.

GARCÍA-CASTRO, Laura. 2019. "Synchronic Variability in the Complementation Profile of *Remember*: Finite vs. Non-finite Clauses in Indian and British English". *Miscelánea: A Journal of English and American Studies* 59: 137-164.

GARCÍA-CASTRO, Laura. 2020. "Finite and Non-finite Complement Clauses in Postcolonial Englishes". *World Englishes* 39: 411-426. DOI: 10.1111/weng.12481

GIVÓN, Thomas. 1985. "Iconicity, Isomorphism and Non-arbitrary Coding in Syntax". In Haiman, John (ed.) *Iconicity in Syntax*. Amsterdam/Philadelphia: John Benjamins: 187-219. DOI: 10.1075/tsl.6.10giv

GREEN, Clarence. 2017. *Patterns and Development in the English Clause System: A Corpus-based Grammatical Overview*. Singapore: Springer.

GREENBAUM, Sidney. (ed.) 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.

HEYVAERT, Liesbet and Hubert CUYCKENS. 2010. "Finite and Gerundive Complementation in Modern and Present-day English: Semantics, Variation and Change". In Winters, Margaret E., Heli Tissari and Kathryn Allan (eds.) *Historical Cognitive Linguistics*. Berlin: Mouton de Gruyter: 132-159. DOI: 10.1515/9783110226447.132

HOFFMANN, Thomas and Lucia SIEBERS. (eds.) 2009. *World Englishes – Problems, Properties*

and Prospects. Amsterdam: John Benjamins. DOI: 10.1075/veaw.g40

LANGE, Claudia and Sven LEUCKERT. 2020. *Corpus Linguistics for World Englishes: A Guide for Research*. London/New York: Routledge. DOI: 10.4324/9780429489433

LEECH, Geoffrey, Marianne HUNDT, Christian MAIR and Nicholas SMITH. 2009. *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge U.P. DOI: 10.1017/CBO9780511642210

LOUREIRO-PORTO, Lucía. 2017. "ICE vs GloWbE: Big Data and Corpus Compilation". *World Englishes* 36 (3): 448-470. DOI: 10.1111/weng.12281

MAIR, Christian. 2006. "Nonfinite Complement Clauses in the Nineteenth Century: The Case of *Remember*". In Kytö, Merja, Mats Rydén and Erik Smitterberg (eds.) *Nineteenth-century English: Stability and Change*. Cambridge: Cambridge U.P.: 215-228.

MAIR, Christian. 2013. "The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars". *English World-Wide* 34 (3): 253-278. DOI: 10.1075/eww.34.3.01mai

ORTEGA, Lourdes. 2003. "Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing". *Applied Linguistics* 24 (4): 492-518. DOI: 10.1093/applin/24.4.492

QUIRK, Randolph, Sidney GREENBAUM, Geoffrey LEECH and Jan SVARTVIK. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

ROHDENBURG, Günter. 1995. "On the Replacement of Finite Complement Clauses by Infinitives in English". *English Studies* 4: 367-388. DOI: 10.1080/00138389508598980

ROHDENBURG, Günter. 1996. "Cognitive Complexity and Increased Grammatical Explicitness in English". *Cognitive Linguistics* 7 (2): 149-182. DOI: 10.1515/cogl.1996.7.2.149

ROHDENBURG, Günter. 2006. "The Role of Functional Constraints in the Evolution of the English Complementation System". In Dalton-Puffer, Christiane, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendl (eds.) *Syntax, Style*

and Grammatical Norms: English from 1500-2000. Bern: Peter Lang: 143-166.

ROHDENBURG, Günter. 2009. "Nominal Complements". In Rohdenburg, Günter and Julia Schlüter (eds.): 194-211. DOI: 10.1017/CBO9780511551970.011

ROHDENBURG, Günter. 2015. "The Embedded Negation Constraint and the Choice between More or Less Explicit Clausal Structures in English". In Höglund, Mikko, Paul Rickman, Juhani Rudanko and Jukka Havu (eds.) *Perspectives on Complementation: Structure, Variation and Boundaries*. Houndmills: Palgrave Macmillan: 101-127. DOI: 10.1057/97811 37450067_6

ROHDENBURG, Günter and Julia SCHLÜTER. (eds.) 2009. *One Language, Two Grammars? Differences between British and American English*. Cambridge: Cambridge U.P. DOI: 10.1017/CBO9780511551970

ROHDENBURG, Günter and Julia SCHLÜTER. 2009. "Introduction". In Rohdenburg, Günter and Julia Schlüter (eds.): 1-12. DOI: 10.1017/CBO9780511551970.001

ROMASANTA, Raquel P. 2017. "Contact-induced Variation in Clausal Verb Complementation: The Case of *Regret* in World Englishes". *Alicante Journal of English Studies* 30: 121-147. DOI: 10.14198/raei.2017.30.05

ROMASANTA, Raquel P. 2019. "Variability in Verb Complementation: Determinants of Grammatical Variation in Indigenized L2 Varieties of English". In Parviainen, Hanna, Mark Kaunisto and Päivi Pahta (eds.) *Corpus Approaches into World Englishes and Language Contrasts*. Helsinki: Varieng. <https://varieng.helsinki.fi/series/volumes/20/romasanta/>. Accessed February 10, 2021.

ROMASANTA, Raquel P. 2021a. "Substrate Language Influence in Postcolonial Asian Englishes and the Role of Transfer in the Complementation System". *English Studies* 102 (8): 1151-1170. DOI: 10.1080/0013838X.2021.1966968

ROMASANTA, Raquel P. 2021b. "A Morphosyntactic Approach to Language Contact in African Varieties of English". *Studia Neophilologica*: 1-24. DOI: 10.1080/00393274.2021.1978314

ROMASANTA, Raquel P. 2021c. "Negation as a Predictor of Clausal Complement Choice in World Englishes". *English Language and*

57

Linguistics: 1-23. DOI: 10.1017/S1360674321000241

RUDANKO, Juhani. 2011. *Changes in Complementation in British and American English: Corpus-based Studies of Non-finite Complements in Recent English*. Houndmills: Palgrave Macmillan.

SCHNEIDER, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge U.P. DOI: 10.1017/CBO9780511618901

STEGER, Maria and Edgar W. SCHNEIDER. 2012. "Complexity as a Function of Iconicity: The Case of Complement Clause Constructions in New Englishes". In Kortmann, Bernd and Benedikt Szmrecsanyi (eds.) *Linguistic*

Complexity: Second Language Acquisition, Indeginization, Contact. Berlin: Mouton de Gruyter: 156-191. DOI: 10.1515/9783110229226.156

VOSBERG, Uwe. 2009. "Non-finite Complements". In Rohdenburg, Günter and Julia Schlüter (eds.): 212-227. DOI: 10.1017/CBO978 0511551970.012

WILLIAMS, Jeffrey P., Edgar W. SCHNEIDER, Peter TRUDGILL and Daniel SCHREIER. (eds.) 2015. *Further Studies in Lesser-known Varieties of English*. Cambridge: Cambridge U.P. DOI: 10.1017/CBO9781139108652

58