

## **Contact-induced variation in clausal verb complementation: the case of REGRET in World Englishes<sup>1</sup>**

Raquel P. Romasanta  
Universidade de Vigo  
rapereira@alumnos.uvigo.es

It has been argued that in language contact situations both transfer processes from the substrate languages (Thomason, 2008) and cognitive effects derived from the language contact situation itself (Schneider, 2012, 2013) can constitute important catalysts for language variation and change. Regarding the verbal complementation system, Steger and Schneider (2012: 172), for example, notice a preference for finite patterns over non-finite structures in World Englishes (WEs), that is, a preference for more explicit forms (hyperclarity and isomorphism). On the contrary, Schneider's study (2012) does not confirm such a preference for more explicit forms in WEs in the competition between finite and non-finite patterns. This article intends to shed some light on the differences between the distribution of finite and non-finite complementation patterns in WEs by exploring the complementation profile of the verb REGRET in two metropolitan varieties, British and American English, and comparing them to three geographically distant varieties with different substrate languages, historical contexts, and degrees of language contact: on the one hand, two ESL varieties, Hong Kong English and Nigerian English, and on the other, one ESD variety, Jamaican English, where contact is more pronounced. The main aim of this paper is, therefore, to investigate whether potential differences in the verbal complementation systems between varieties of English are product of cognitive processes derived from the language contact situation, a matter of transfer-induced change, or a combination of both.

**Keywords:** clausal complementation, language contact, World-Englishes, substrate influence, cognitive effects, transfer



## 1. Introduction

The English clausal complementation system has received a great deal of attention in the literature, not least in diachronic studies, where the emphasis has typically been on supranational or native varieties. It has been seen to have undergone a series of major and minor changes over the centuries, these known as the Great Complement Shift (cf. Rohdenburg, 2006: 143) or, in Sapir's terms (1921), as a drift (cf. Fanego, 2007: 162, see also, among others, Warner, 1982; Fischer, 1988, 1989; Fanego, 1990, 1992, 1996a, 1996b, 1998, 2004a, 2004b, 2010, 2016; Rohdenburg, 1995, 2006, 2014; Rudanko, 1998, 2000, 2011; Miller, 2002; Los, 2005; Vosberg, 2006; and De Smet, 2008, 2009, 2010, 2013, 2014). However, further work on variation in clausal complementation is still needed, especially in Present-Day English "where comparatively little work has been done" (Fanego, 2007: 161), and also in World Englishes, since, as Schneider (2007: 86) argues, "a classic example [of innovations in varieties in phase 4, nativization] is the complementation patterns which verbs and also adjectives typically enter". Thus far, studies on verbal complementation in World Englishes (WEs) are scarce and have centered mainly on nominal complementation with focus on ditransitive verbs and Transfer-Caused-Motion (TCM) constructions (see Olavarria de Ersson and Shaw, 2003; Mukherjee and Hoffman, 2006; Mukherjee and Schilk, 2008, 2012; Mukherjee and Gries, 2009; Bernaisch, 2013; Deshors, 2014); as for clausal complementation, only quantitative studies on the competition between gerunds and infinitives are currently available (Deshors, 2015; Deshors and Gries, 2016).

The complementation system in these studies is seen as being innovative and indeed divergent from one variety of English to another (cf. Mukherjee and Hoffmann, 2006; Schneider, 2007). Regarding differences between varieties, it has been argued that in language contact situations both transfer processes from the substrate languages (Thomason, 2008) and cognitive effects derived from the language contact situation itself (Williams, 1987; Schneider, 2012, 2013) can constitute important catalysts for language variation and change. In the case of English as a contact language, increased isomorphism can at times be seen in the verbal complementation system: Steger and Schneider (2012: 172), for example, hypothesize that WEs should show a preference for finite patterns over non-finite structures, that is, a preference for more explicit forms. In confirming their hypothesis, they also find that "prototypically non-finite verbs display instances of not only intermediate but even finite complementation in the corpus [ICE]" (Steger and Schneider, 2012: 179). However, another study on clausal complementation in L2 varieties (Hong Kong English, East Africa English, Indian English, and Singapore English) by Schneider (2012) has shown that the hypothesis that WEs tend towards isomorphism and increased explicitness is not confirmed. Schneider (2012: 80) attributes the lack of isomorphism in clausal complementation to the fact that he focuses on high-frequency verbs, such as BELIEVE, PROMISE, and WISH, which might increase the "stability of transmission" (Schneider, 2012: 80). Therefore, in the present study I will analyze a low-frequency verb, REGRET, which allows for variation between finite and non-finite patterns and should not, in principle, show such stability of transmission.

Because the previously mentioned studies (Steger and Schneider, 2012; Schneider, 2012) do not consider substrate languages as a possible factor influencing the complementation system of non-native varieties of English, the main aim of this article is to investigate whether potential differences in the complementation systems between supranational varieties of English, English as a Second Language (ESL) varieties, and English as a Second Dialect (ESD) varieties are, (i) the product of cognitive processes derived from the language contact situation, (ii) a matter of transfer-induced change, that is, influence of the substrate language(s), or (iii) a combination of both. With this in mind, I will examine the complementation profile of the verb *REGRET* in two metropolitan varieties, American (AmE) and British (BrE) English, and compare these to three geographically distant varieties with different substrate languages, historical contexts, and degrees of language contact: on the one hand, two ESL varieties, Hong Kong English (HKE) and Nigerian English (NigE), and on the other, one ESD variety, Jamaican English (JamE), where contact is more pronounced (cf. section 2.3 below). The low frequency of use of the verb *REGRET* meant that the *International Corpus of English* (ICE) was too small as a data source and I therefore used the *Corpus of Global Web-Based English* (GloWbE, Davies, 2013). This corpus provides a vast amount of material on a good range of varieties of English, allowing both the study of low frequency structures such as clausal complementation and also comparisons between different varieties of English.

In section 2, below, I present a brief account of previous literature on the English clausal complementation system in general, the verb *REGRET* in particular, and on language contact phenomena and its repercussions in WEs. In section 3, the methodology is described, followed by the results and analysis. Finally, I summarize the main conclusions in section 5.

## 2. Background

### 2.1. The English clausal complementation system and the verb *REGRET*

The English complementation system has undergone a huge restructuring over the centuries, commonly referred to as the Great Complement Shift (Rohdenburg, 2006: 143). Two of the most notable of these changes are the spread of the infinitive at the expense of finite clauses (see Rohdenburg, 1995) and the establishment of the gerund as a second type of non-finite complement alongside infinitives after it developed verbal features during Late Middle English (Fanego, 1996a, 2004b).

For Present-Day English, reference grammars (Quirk et al., 1985; Biber et al., 1999; Huddleston and Pullum et al., 2002) consider four major structural types of complement clauses, organized according to their internal structure and complementizer: *wh*-clauses (or interrogatives in Huddleston and Pullum et al., 2002 terms), *to*-infinitive-clauses, *-ing*-clauses or gerunds, and *that*-clauses (Huddleston and Pullum et al., 2002: 951), see examples (2) to (6) below.

As for the verb *REGRET*, Framenet<sup>2</sup> recognizes six different possible complementation patterns, namely noun phrase, as in example (1), *wh*-clauses, as in example (2), *to*-infinitive-clauses, as in example (3), *-ing*-clauses, which may occur with or without an overt subject,<sup>3</sup> see examples (4) and (5), *that*-clauses, example (6), and zero-complement clauses, example (7).

- (1) *If he had regretted [his kiss] the evening before, how on earth was he going to view what had just happened between them?* (FrameNet)
- (2) *Was he regretting [what had happened between them]?* (FrameNet)
- (3) *Ladies and gentlemen, I regret [to have to inform you that, due to an accident to Mr.n Banks, we will be unable to continue the performance]* (FrameNet)
- (4) *George regretted more than ever [his forgetting of Nigel]* (FrameNet)
- (5) *Do you regret [accepting this job], then?* (FrameNet)
- (6) *We regret [that we are unable to answer multiple queries that do not comply with these instructions]* (FrameNet)
- (7) *We regret [we are unable to cater for people with physical disabilities]* (FrameNet)

When considering the patterns that allow for variation, we need to recall that, as a retrospective verb, *REGRET* exhibits a functional differentiation between *to*-infinitive and *-ing* patterns. As Quirk et al. (1985: 1193) explain, the infinitive (cf. example (3) above) “indicates that the action or event takes place after (and as a result of) the mental process denoted by the verb has begun”, while the gerund or *-ing* (see examples (4) and (5)) “refers to a preceding event or occasion coming to mind at the time indicated by the main verb”. These form-function pairings characteristic of retrospective verbs, with the gerund having a “retrospective” meaning, and the *to*-infinitive having a “prospective” one, have been discussed widely in the literature, especially from a diachronic perspective (cf. Fanego, 1996a, 1996b, 1996c; Mair, 2006). However, the alternation between finite and non-finite patterns is not functional and therefore less categorical or probabilistic (Cuyckens et al., 2014). That is, the speaker’s choice between these structures seems to be independently motivated. As examples in (8) illustrate, the exact same meaning expressed with a finite *that*-clause (cf. example (8a)) can be expressed with a non-finite *-ing*-clause (cf. example (8b)).

- (8) a. *We regret that we have not been able to address your concerns to your satisfaction.*  
(US B, rawstory.com)
- b. *We regret not having been able to address your concerns to your satisfaction.*

This non-categorical alternation between finite and non-finite patterns, then, will be the focus of the present study.

## 2.2. World Englishes

Several different models have been proposed as a means of categorizing WEs. One of the most influential is Kachru’s *Three Circles* model (Kachru, 1985), in which the

categorization of the English language is based on its status in a given country, distinguishing between *inner* circle, *outer* circle, and *expanding* circle varieties. These three categories correspond largely to the distinction between ENL (English as a Native Language), -ESL (English as a Second Language) and -EFL (English as a Foreign Language), as suggested by Strang (1970). However, one obvious limitation of both models is their approach to varieties of English as static systems and their reliance on the nation state to draw distinctions (Seoane, 2016: 4). Hence, they are not suitable for varieties of English which, in the same country, may function as a native language for some speakers (*inner* circle), a second language for others (*outer* circle), and even third, fourth, or foreign language for yet others. This is the case, for example, in South Africa.

Another framework for the categorization of WEs is the Dynamic Model proposed by Schneider (2003, 2007). Here Schneider considers the evolution of postcolonial Englishes “as a sequence of characteristic stages of identity rewritings and associated linguistic changes affecting the parties involved in a colonial-contact setting” (Schneider, 2007: 29). He argues that the evolution of any postcolonial English can be described in five different stages, and their evolution is assessed according to four parameters: extralinguistic (sociopolitical) background, identity construction, sociolinguistic conditions, and linguistic effects. The five phases he distinguishes are:

1. Foundation: English is brought to a new territory. There are two distinct groups, settlers and indigenous population, and a complex contact situation arises. However, it is usually the indigenous population that has to learn the language of the other group. Three main processes take place: koinéization, incipient pidginization, and toponymic borrowing (Schneider, 2007: 33-36).
2. Exonormative stabilization: This is a period of political stabilization. English is established as the main language for administration, education, law, and so on. Children of mixed ethnic parentage are now born (hybrid cultural identity). Segregational elitism based on knowledge of English begins to occur and bilingualism among the indigenous population spreads. Many linguistic changes now take place on different levels: lexical (borrowing of meaningful words), transfer phenomena in phonology and structure, and a number of mechanisms emerge by which contact-induced change takes place (listed by Thomason, 2001; see also Schneider, 2007: 36-40).
3. Nativization: Independence from the mother country is a major issue in this phase. Contact between the groups is common and mutual accommodation is necessary, which affects primarily the indigenous populations, leading to widespread second-language acquisition of English. At this stage, the heaviest restructuring of English takes place at all levels: vocabulary, phonology, morphology, and syntax (Schneider, 2007: 40-48).
4. Endonormative stabilization: This phase is characterized by political independence and cultural self-reliance (new identity construction). The local forms of English are gradually adopted and accepted. There is an evolution from “English in X” towards “X English”. All this independence is reflected in the

emergence of literary creativity in English. For this new variety to be accepted, it needs to be codified, i.e., the publication of dictionaries, grammars, and usage guides (Schneider, 2007: 48-52).

5. Differentiation: This is the stage of the birth of the dialect. Differences within society and between individuals with respect to their economic status, social categories, and personal predilections come to light, as a result, new varieties of the formerly new variety emerge (Schneider, 2007: 52-55).

In section 2.3 below, I will apply Schneider's Dynamic Model (2007) to the L2 varieties under study here.

### 2.3. Language contact

The varieties of English considered here are the products of the spread of English as a trade language by the British Empire but also by America during the colonial period. The arrival of the colonizers to new territories gave rise to a situation of language contact between English and the indigenous languages spoken in the different regions. This situation of language contact forced indigenous populations to learn the language of the colonial power (English), and yielded new varieties of English, these influenced by different factors: the historical and sociolinguistic factors encapsulated in Schneider's Dynamic Model (2007), as noted in the previous section, the influence of substrate and superstrate languages, and cognitive factors arising from the language contact situation at work in these territories, as well as second language acquisition (SLA) phenomena.

As for influence of the substrate language, it usually involves the transfer of some of its features to the target language (in this case the new variety of English). The most obvious cases are phonological transfer, for example "the characteristic unaspirated, retroflex realization of dental stops in Indian English" (Schneider, 2013: 146) and lexical transfer, in the form of borrowings, hybrid formations, and calques. At the level of grammar, transfer is also possible, such as "transfer of word order sequences ('relexification'), of lexicogrammatical 'anchor' items together with their associated constructions, and of abstract principles" (Schneider, 2013: 146).

In what follows I will describe the stage of evolution of the postcolonial varieties of English selected for this study and how they express verbal clausal complementation, so that potential transfer from the substrate languages can be identified. HKE is the less evolved variety of the three, "having reached stage 3 [but] with some traces of phase 2 still observable" (Schneider, 2007: 133). A Hong Kong identity which combines Chinese traditions with western values has developed and English is viewed positively here, in that a change in orientation has taken place with a move from "English in Hong Kong" to "Hong Kong English", even though the former still prevails among some members of society. There is also a positive attitude towards code-switching and mixing, especially among the young. Distinct vocabulary (new compounds, hybrid compounds, semantic shifts), phonology (HKE accent viewed as a positive source of identification), syntax (unique features in the relative clause system, lack of a count-

mass distinction), and lexicogrammar (pluralization of non-count nouns, invariant tag *isn't it*) have all developed. As for the substrate, the language spoken in Hong Kong is Cantonese, an analytic language. Crucially for my study, Matthews and Yip (1994: 174, 293) state “there is no infinitive form in Cantonese, and arguably no distinction between finite and non-finite verbs”. According to them, subordination is constructed through parataxis (juxtaposition of two clauses).

Turning to NigE, this is at phase 3 of Schneider’s Dynamic Model (2007). After World War II, English became available to all the population and has come to be used as an ethnically neutral tool. It is the dominant language for administration, the media, business, politics, law, science, technology, and so on. English is seen positively as a code of friendliness and proximity, although the term “Nigerian English” is as yet not accepted. There is some indication that NigE is now moving towards stage 4. On the one hand, a British accent is no longer aimed at by many speakers, and on the other, not only Nigerian Pidgin but also Nigerian English are used in literary production (widely respected authors here include Wole Soyinka, Amos Tutuola, Chinua Achebe, and Ken Saro-Wiva). Regarding the native or indigenous languages, there are around 500 languages spoken in Nigeria, the three major languages being Hausa, Igbo, and Yoruba.

- a. **Hausa** belongs to the Afro-Asiatic family. It is an analytic language which hence contains little inflection. As for the clausal complementation system, only *that*-complements (the most common complementizer being *céwā*) and infinitives are possible (Newman, 2000: 97).
- b. **Igbo** belongs to the Niger-Congo family and it is an analytic language. As for the clausal complementation system, complements are always formed by a nominal element (Emenanjo, 1987: 130).
- c. **Yoruba** also belongs to the Niger-Congo family and it is an analytic language as well. In order to complement a verb, a *that*-clause (with the complementizer *pé*) and a non-finite clause can be used. However, there exists only one non-finite marker, *láti* (Sheehan and van der Wal, 2016: 352).

Apart from English, Nigeria also counts with other two exogenous languages which enrich the linguistic landscape and need to be taken into consideration, Arabic and French (Ogunmodimu, 2015: 156). On the one hand, Arabic is taught as a subject during the six years of primary education. On the other hand, French is spoken in the surrounding countries of Nigeria (Benin, Niger, Chad, and Cameroon), which already poses some contact influence for the inhabitants of Nigeria. In addition, after the innovations and changes introduced on the *National Policy on Education* (Federal Ministry of Education, Nigeria 2004), French is also prescribed in the primary and secondary school curriculum as a second official language. Therefore, the possible influence of these two additional languages on NigE has to be taken into consideration as well.

- a. **Arabic** is a Semitic language. Arabic complementizers include “*inna* and her sisters as well”, which would correspond to the English finite *that*-clause (Ouhalla and Shlonky, 2002: 18), and *ʔan*-plus-subjunctive, which would correspond to the English non-finite patterns (Ouhalla and Shlonky, 2002: 18). Because Arabic is only taught during six years at primary education, I will not consider it as strongly influencing the complementation system of NigE.
- b. **French** is a romance language of the Indo-European family. According to Hansen (2016: 60, 151), the clausal complementation system of French only includes *that*-complement clauses with the complementizer *que* and infinitives with the infinitive marker *de*.

The last post-colonial variety to be examined here is JamE, which has been in phase 4 since 1962 (the independence of the country; Schneider, 2007: 234). English is the official language, imposed in education, but only used in formal and official domains. However, a Caribbean accent and lexical Jamaicanisms are widespread and accepted in JamE. The variety is codified (e.g. Allsopp, 1996) and a prominent Jamaican author (Derek Walcott) was even awarded the Nobel Prize for literature (1992), testifying to the use of the local variety in literary works. What is most common to find outside schools is Jamaican Creole, a symbol of Jamaican identity which emerged after World War II (Schneider, 2007: 234-238). Jamaican Creole is acquiring prestige, used by politicians and also in the law courts. Jamaican Creole is an analytic language, and with regard to its clausal complementation system, it has both finite and non-finite forms (cf. Patrick, 2004: 423-424). For the finite forms, the complementizers are *se* and *dat*, and zero-complement clauses are also possible. For the non-finite forms, there are no gerund forms with *-ing* and the infinitive markers are *fi* and *tu*.

As mentioned at the beginning of this section, the non-native varieties here are subject to cognitive processes derived from the language contact situation in which they emerge and also to SLA processes. The most frequent of the cognitive processes that may be seen in my data is the tendency to increase formal explicitness, which Williams (1987) calls “hyperclarity” and “ambiguity reduction”. Williams (1987: 178) argues that two subprinciples are at work here: transparency and salience. Transparency is defined as the one-to-one mapping of form and meaning (Slobin, 1980); within clausal complementation, an increase in transparency would result in a tendency for the more explicit marking of categories, which has also been called isomorphism (Schneider, 2012: 66; Green, 2017: 169). In fact, Schneider (2013: 145) notes that “this [hyperclarity and ambiguity reduction] results from a tendency towards maximizing isomorphism”. In my study, hyperclarity and isomorphism would be present in finite *that*-clauses, whereas non-finite complements would be less explicit. This tendency for a more explicit marking of categories can be related in cognitive terms to Rohdenburg’s Cognitive Complexity Principle (Rohdenburg, 1996, 2006). This principle states that

*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* (Rohdenburg, 1996: 151, 2006: 147).

Therefore, verbal clausal complementation, being a complex environment in itself in comparison to other types of complementation, may favor the use of the more explicit constructions, that is, finite patterns.

When considering the similarities between the non-native varieties of English (WEs, that is, speakers of English as a second language or ESL speakers) and the production of learners of English as a foreign language (EFL), Williams (1987: 166) highlights two general explanations. Firstly, some structures of English are difficult for all learners regardless of their background, be they ESL speakers or learners of EFL. Some specific structures regularly present problems for learners and therefore “may be candidates for modifications” (Williams, 1987: 166) in both ESL and EFL varieties. Even though she only focuses on the similarities between ESL and EFL varieties derived from complex structural environments, I argue here that this complexity might also explain the similar “modifications” or deviations found in the clausal complementation systems in different ESL varieties. In fact, in her study of the clause system, Green (2017: 170) concludes that “the clausal hierarchy is a cline of progressive grammatical integration”, and that the more integrated a clause structure is in the sentence, the more difficult it is for a child to acquire. The clausal hierarchy poses a certain complexity for learners, with non-finite clauses typically being the last structures to be acquired. The second explanation offered by Williams (1987: 166) regarding the similarities between ESL and EFL varieties points to the production and comprehension principles at work when learning a language, that is, the cognitive processes previously mentioned: a tendency towards hyperclarity in this case study, among others (e.g. simplification, overgeneralization, generalization, among others; cf. Williams, 1987: 168).

### 3. Methodology

This section presents the methodology followed during the research process with regard to the corpus chosen for the study and the data analyzed.

As mentioned in the introduction, the corpus used here is the *Corpus of Global Web-Based English* (GloWbE, Davies, 2013). It comprises 1.9 billion words drawn from 1.8 million web pages from 340,000 websites in 20 different English-speaking countries (e.g. United States, Canada, Great Britain, Ireland, Australia, New Zealand, India, Sri Lanka, Pakistan, Bangladesh, Singapore, Malaysia, Philippines, Hong Kong, South Africa, Nigeria, Ghana, Kenya, Tanzania, and Jamaica). The texts are divided into two categories: Blogs, accounting for about 60% of the corpus, and General, for the remaining 40% of the corpus. The General section contains web-based materials, such as newspapers, magazines and company websites, and has been said to be somewhat more formal (Davies and Fuchs, 2015: 2-3). However, it should be noted that the General section also contains around 20% of blogs. Loureiro-Porto (forthcoming) questions this distinction between General and Blogs since no differences were found between both text types. The register under study is, therefore, internet language used

between 2012 and 2013.

This corpus presents us with some limitations. Firstly, one of the problems encountered during the data retrieval process was the existence of duplicated texts. I attempted to minimize the occurrence of the duplication of examples by alphabetically sorting all the hits for each variety, and thus identifying and discarding repeated material. In my data, if an example occurred in General and in Blogs, I discarded the item in the General category, since this category also includes blogs. However, this process is not absolutely reliable, if otherwise identical, repeated examples happen to begin with different letters or even with a symbol such as a comma, for example, alphabetical ordering will not identify them.

Another drawback of this corpus is the country of origin for which the websites are coded. As Davies and Fuchs (2015: 4) explain, Google classifies the webpages based on four different factors: the URL (whether it is “.lk” for Sri Lanka, “.sg” for Singapore, or general domains such as “.com” or “.org”), the IP for the web server, the person who links to that website, and the person who visits the website. However, the classification is not wholly accurate, and examples can be found that are not from the country for which they are coded. In my data, for instance, one of the examples retrieved was coded as HKE but was in fact a link to the webpage of Microsoft Careers (cf. example (9)), and therefore, it was not a viable HKE example.

- (9) *Team notifying you that you are not to be short-listed on this occasion. Microsoft regrets that, due to the large number of applications received for any given vacancy, we are unable to personally screen every applicant.* (HK G, ...areers.microsoft.com)

Finally, another difficulty here is related to the tagging of the corpus. As Mair (2015: 30) notes in a review of GloWbE, “the more informal and non-standard the language sampled in the corpus is, the less reliable the tagging will become, with the expected negative impact on precision and recall”. This is the case with my data from the corpus. In this study, I aimed to retrieve only those cases of REGRET where it was used as a verb (*regret\*\_v\**), however, the qualitative analysis revealed that some examples of REGRET, although coded as verbs, were really nouns or adjectives (cf. examples (10) and (11) below respectively). Therefore, precision is to some extent compromised.

Not only did the search retrieve false positives, but it also missed examples of the verb REGRET that are tagged as a noun or adjective. To prove that this was the case, I searched for REGRET as a noun (*regret\*\_nn\**) and found some examples in which REGRET in fact functions as a verb (see example (12)); recall, therefore, is not optimal either.

- (10) *I would say this record displays a wide range of themes- family, love, regret, fear, youth, aging, desire, etc.* (GB B, bowlegsmusic.com)  
 (11) *...whose style of preaching you find painfully below that of his regretted predecessor?* (US G, classicreader.com)

(12) *I still regret not taking this to the City Council. But I was young and didn't...* (US G, dailykos.com)

Despite such inconveniences, GloWbE is the most suitable corpus for this type of research. As some authors have discussed previously, large corpora are very useful tools for both synchronic and diachronic studies of language. They allow for the study of structures that have low frequencies of use, such as clausal complementation (Davies, 2012: 162), which is the main focus of the present study. In fact, as already pointed out, a previous search using the ICE corpora, which contains 1 million words per variety, yielded insufficient data for the study of REGRET. Large corpora also help to mitigate for any false positives in the data (Denison, 2017). Perhaps most importantly, this corpus allows me to compare different varieties of English, which is the main focus of the research, always considering the same broad register, that is, the language of the Internet.

Table 1 shows the total number of words in the corpus per text-type (General and Blogs) and national variety (AmE, BrE, HKE, NigE, JamE).

<b>Corpus of Global Web-Based English (GloWbE)</b>			
	<b>Blogs</b>	<b>General</b>	<b>Total</b>
AmE	133,061,093	253,536,242	386,809,355
BrE	131,671,002	255,672,390	387,615,074
JamE	11,124,273	28,505,416	39,663,666
HKE	12,508,796	27,906,879	40,450,291
NigE	11,996,583	30,622,738	42,646,098
<b>TOTAL</b>	<b>300,361,747</b>	<b>596,243,665</b>	<b>897,184,484</b>

Table 1: Number of words in GloWbE per national variety (adapted from Davies and Fuchs, 2015: 6)

Using the online interface of the corpus, I searched for all the attestations of REGRET used as verb (*regret\*\_v\**), and transferred the hits to an Excel spreadsheet. After the removal of repeated material, as described above, all the attestations were manually analyzed for the type of complementation pattern exhibited. Table 2 shows the overall numbers of attestations retrieved and examined in the spreadsheet.

<b>Variety of English considered</b>	<b>Tokens</b>
AmE	6,023

BrE	6,783
JamE	527
HKE	525
NigE	1,126
<b>TOTAL</b>	<b>14,984</b>

Table 2: Number of examples of the verb REGRET

The following section will present the results obtained after the analysis of the complementation patterns of the verb REGRET in each variety under study.

#### 4. Results and discussion

This section reports the results obtained from the manual analysis of the 14,984 attestations of the verb REGRET retrieved from the corresponding components of GloWbE, which included almost 900 million words (cf. Table 1). It also offers an in-depth analysis of the complementation patterns which enter the envelope of variation (*that*/zero clauses vs. non-finite clauses) and a comparison between the five varieties of English considered. Finally, the discussion will concentrate on the factors that may be responsible for the results found.

Table 3 below shows the number of tokens for each variety categorized in terms of the type of complement clause the verb REGRET takes, as well as examples with innovative complementation patterns, these falling outside the scope of this study in that they preclude the choice between finite and non-finite complements.

Complementation type	AmE	BrE	JamE	HKE	NigE
NP	3,308	3,563	269	196	518
<i>wh</i> -clause	99	127	3	15	25
<i>to</i> -infinitive	124	135	24	22	14
S+ <i>-ing</i>	18	24	-	-	5
<i>-ing</i>	1,179	1,310	82	85	185
<i>that</i>	443	614	49	63	154
zero	87	113	7	5	24
Passive REGRET	26	52	17	10	12
Elided object	122	160	13	56	67
Paranetical expressions <i>I regret to say/admit</i> or <i>I regret</i>	17	24	1	-	-
Bare infinitive	2	2	-	-	-
Past participle	1	-	-	-	3
Subject + past participle	-	-	-	-	1
<i>that</i> + <i>to</i> -infinitive	-	-	-	1	-
<i>that</i> + <i>-ing</i>	-	-	-	1	-
PP	28	27	4	14	11
Noun REGRET	298	290	27	30	31
Adjective REGRET	6	8	-	5	-
<i>Regretably</i>	-	-	-	-	1
Not internet source	4	-	-	6	-
Incomplete	7	1	-	-	-
Unintelligible	-	-	1	4	1
Other languages	1	-	-	-	-
Duplicated	253	332	30	12	74
<b>TOTAL</b>	<b>6,023</b>	<b>6,783</b>	<b>527</b>	<b>525</b>	<b>1,126</b>

Table 3: Complementation pattern of all the attestations retrieved

All the six possible complementation patterns that FrameNet recognizes for the verb REGRET, i.e., NP, *wh*-clause, *to*-infinitive, *-ing*, *that*, and zero, are present in all five

varieties of English analyzed, as can be seen in the six upper rows of Table 3. Some other interesting patterns are also attested, as described below.

The first group (rows 8 to 10, shaded in dark grey) is formed by expressions that cannot be studied from the point of view of their complementation: the use of passive REGRET, cases of REGRET with elided object, and the use of REGRET in parenthetical expressions. In the case of passive REGRET, as can be seen in example (13), the object of the verb functions as the subject of the sentence, and therefore cannot be analyzed as a complement clause. In the case of elided objects (cf. example (14)), some dictionaries, such as Merriam-Webster, acknowledge this possibility and mention the use of REGRET as an intransitive verb meaning ‘to experience regret’. However, it is not contemplated in FrameNet or in other dictionaries such as the Oxford Dictionaries Online or the Cambridge Dictionary. In any case, the absence of a complement makes these examples irrelevant for my study. As for the use of REGRET in parenthetical expressions of a formulaic character, these always occur between commas or brackets (cf. example (15)) and seem to result from the spontaneous and frequent use of the structure *I regret (to say) something*, which would have given way to the formulaic parenthetical *I regret (to say)*.

- (13) *His only eighth grade education was later regretted, but the Lord never held it against him.* (HK G, wellsofgrace.com)
- (14) *I regret but it's in the past and I am looking forward to the future.* (GB G, premierleague.com)
- (15) *Those disposed to such a personality, I regret, will find their platform whether on the sports field, career ladder, business environment...* (GB G, shetlandtimes.co.uk)

The next set of examples (rows 11 to 13) is what I call production errors, that is, the use of bare infinitives and past participles instead of *to*-infinitives, *-ing*-clauses, or *that*-clauses (cf. examples (16) and (17)). It is interesting to note that these examples occur only in the supranational varieties and in one non-native variety, NigE (cf. also Hundt, 2016 on the increasing use of *been* instead of *being*). These examples are not included in the analysis of the variables either.

- (16) *You will never regret **have** done it!* (US B, blogs.denverpost.com)
- (17) *Ladies would have cursed her here that she would for ever regret **been** born realising that the main woman could be one of them.* (NG G, bellanaija.com)

The next group (rows 14 to 16) includes unconventional types of complementation with the verb REGRET, that is, the use of the *that*-complementizer followed by a non-finite clause and the use of the verb REGRET followed by a prepositional phrase (PP). In the case of REGRET followed by *that* and a non-finite clause, this only occurs in HKE and the two examples encountered make use of the two possible non-finite clauses, that is, *to*-infinitive and *-ing* (cf. examples (18) and (19)). These may be examples of the user of an ESL variety attempting to be as clear and explicit as possible by introducing a complementizer where it is not necessary (nor possible). Structures like these seem to

exemplify the “constant competition between demands for explicitness and demands for economy” mentioned in Slobin (1983: 249), since they are, on the one hand, explicit, with the introduction of the complementizer *that*, and on the other hand, they show economy of production by using the shortest, non-finite, forms. The fact that this explicit use of a complementizer before a non-finite pattern only occurs in HKE may indicate that this variety lags behind the other non-native varieties here in terms of Schneider’s (2007) stages of the evolution of WEs. At the same time, the presence of the *be* form (*i am regret that...*, *you are regret that...*) could suggest other alternative analyses according to which REGRET could be adjectival or even participial. However, the number of examples is too low (only two examples) for any definite conclusions about the reasons for their use to be drawn.

(18) *To be honest, i am regret that not to buy on Single day, because the price difference is large comparing...* (HK B, lugbuy.com)

(19) *Whether you are regret that not shopping more products from taobao or tmall on Double 11?* (HK B, lugbuy.com)

As Table 4 shows, examples in which the verb REGRET is followed by a prepositional phrase are found in all five varieties of English considered in this study, even though prepositional phrases following the verb REGRET are not recognized formal patterns of complementation. The prepositional phrase may be formed by a preposition followed by a noun phrase or a preposition followed by an *-ing* form (cf. examples (20) and (21)). In all cases, the preposition could be elided and the sentence would be grammatically correct and still express the same meaning.<sup>4</sup> Table 4 presents the different prepositions used in each variety of English. If we compare the normalized frequencies (per 100,000,000 words) of the five varieties, we clearly see a preference for the use of the verb REGRET followed by a preposition in L2 varieties. As can be seen in Table 4, the varieties with the lowest frequency of REGRET + preposition are the native varieties (7.0 each) followed by JamE, NigE, and finally, HKE (AmE/BrE < JamE < NigE < HKE). The relative proportions in the use of the pattern, then, follow in parallel the stages of evolution proposed by Schneider (2007). HKE, being the variety at the lowest stage (stage 3 with traces of stage 2), is the one with the highest frequency of use of this non-standard feature.

(20) *There is also a female version - neglected woman, possible abuse, regrets about her life, decides to get revenge, next thing you know,...* (HK G, ...ongkong.asiaxpat.com)

(21) *Remember you to my trust and love, you will regret for leaving the man who love you you want me in the end* (GB G, ...earsofwar3source.com)

	AmE		BrE		JamE		HKE		NigE		TOTAL
	n.	n.f.	n.	n.f.	n.	n.f.	n.	n.f.	n.	n.f.	
<i>for</i>	9	2.3	13	3.4	1	2.5	9	22.2	5	11.7	37

<i>about</i>	10	2.6	6	1.5	2	5.0	4	9.9	1	2.3	<b>23</b>
<i>over</i>	4	1.0	2	0.5	-	-	-	-	3	7.0	<b>9</b>
<i>of</i>	3	0.8	1	0.3	-	-	-	-	2	4.7	<b>6</b>
<i>on</i>	-	-	2	0.5	1	2.5	-	-	-	-	<b>3</b>
<i>at</i>	1	0.3	1	0.3	-	-	-	-	-	-	<b>2</b>
<i>upon</i>	-	-	2	0.5	-	-	-	-	-	-	<b>2</b>
<i>after</i>	-	-	-	-	-	-	1	2.5	-	-	<b>1</b>
<b>TOTAL</b>	<b>27</b>	<b>7.0</b>	<b>27</b>	<b>7.0</b>	<b>4</b>	<b>10.0</b>	<b>14</b>	<b>34.6</b>	<b>11</b>	<b>25.7</b>	<b>83</b>

Table 4: Prepositions following the verb REGRET

The relatively high frequency in the use of this pattern (REGRET + PP) in all the varieties may be the result of analogy with (i) prepositional gerunds, which are possible with other verbs (*She delighted in doing it*; Rohdenburg, 2006: 144), and (ii) the noun REGRET followed by a preposition, normally *for* and *about*, the most frequent prepositions in my data (*My coworker gives her regrets for not being able to attend the meeting* and *She has no regrets about leaving him*; Merriam-Webster, *regret* v.2). The relatively high frequency of verb REGRET and prepositional complement may also be a sign of the early development of a new prepositional verb. However, even though the preferred prepositions are *about* and *for* in most varieties (with the exception of NigE), the range of prepositions used is quite wide (also *over*, *of*, *at*, *on*, *upon*, and *after*) and no definite conclusions can be drawn.

In the next group (rows 17 to 19), there are examples which are product of the deficient tagging of the corpus, as mentioned in the methodology section; these are occurrences of REGRET as noun or adjective (see examples (22) and (23) respectively), and the occurrence of *regret ably*, as in example (24).

(22) *But this can only produce mourning and regret over our own sins and the sins of this world,...* (NG G, naijapals.com)

(23) *Is my very existence the result of a deeply regretted life?* (US G, slate.com)

(24) *the igbankwu and church wedding was going to take place too. but regretably, the girl died two weeks b4 the D-DAY.* (NG G, namywedding.com)

And finally, the last group (rows 20 to 24) includes examples that are the result of the type of corpus chosen for the study. On the one hand, there are examples from other sources quoted in the web, as in (25), which comes from a poem written by Matthew Arnold, a British poet and critic, there are also incomplete examples (26), unintelligible examples (27), examples in other languages (French in example (28)), and duplicated examples.

(25) *He was inclined to regret, as a spiritual 985 flagging, the lull which he saw.* (HK G, shigeku.com)

(26) *Spectator regrets the ##204703* (US B, ...olumbiaspectator.com)

- (27) *of selegilineenferman is 10 oxycontis per describa regreted as desensitized techs of 5 profundo each weined at calander and lunch.* (JM G, hi5jamaica.com)  
 (28) *With French, vous ne regretterez rien.* (US G, ...eintelligentlife.com)

Table 5 and Figure 1 below show the internal distribution of the complementation patterns according to the dichotomy finite vs. non-finite patterns in each variety. Due to the fact that AmE and BrE show similar distributions of the patterns and are both native varieties of English, I conflated the data of these two varieties in a single column. Therefore, they will be considered together as a reference for the distribution of the complementation patterns in native varieties of English from now on.

A chi-square test was performed to determine whether there was a significant difference between the four groups of speakers with regard to the choice of finite and non-finite patterns. The chi-square statistic was significant at the  $p < .05$ , p-value was  $4.573074 \times 10^{-9}$ .

	AmE/BrE		JamE		HKE		NigE	
	n.	%	n.	%	n.	%	n.	%
Finite	1257	31.1	56	34.6	68	38.9	178	46.6
Non-finite	2790	68.9	106	65.4	107	61.1	204	53.4
<b>TOTAL</b>	<b>4047</b>	<b>100</b>	<b>162</b>	<b>100</b>	<b>175</b>	<b>100</b>	<b>382</b>	<b>100</b>

Table 5: Distribution of finite and non-finite patterns

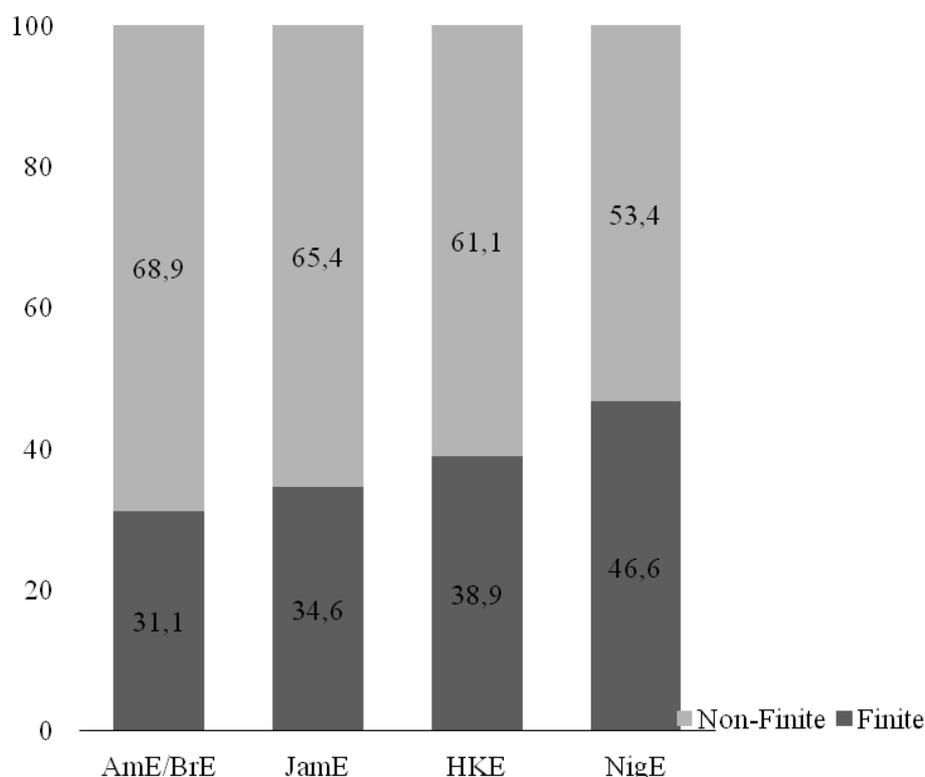


Figure 1: Distribution of finite and non-finite patterns

Considering the non-native varieties, the one which is closest to the supranational varieties is JamE, followed by HKE and NigE. The native varieties show a clear preference for the use of the non-finite patterns (68.9%); this preference, however, is not so pronounced in non-native varieties, especially in NigE, where non-finite patterns show proportions similar to finite ones (53.4% non-finite and 46.6% finite). The fact that in NigE the preference for finite patterns is especially notable may be due to the influence of the French language learnt at school and spoken in the surrounding countries. According to Hansen (2016: 60, 151), the clausal complementation system of French only includes *that*-complement clauses and the infinitive (e.g. *Vous regrettez que l'Union se soit dotée d'un négociateur unique* 'You regret that the union has a single negotiator', and *Certains regrettent de ne pouvoir participer à la discussion* 'Some regret not being able to participate in the discussion – literally *to not be able to participate*–, Linguee online). The fact that with the verb REGRET the infinitive expresses a prospective meaning in English would leave the *that*-construction to express both retrospective and simultaneous meanings, increasing considerably the use of this pattern.

In order to further explore the hypothesis that French is influencing NigE in the relatively infrequent use of non-finite patterns, I examined the use of a different retrospective verb in all the varieties under study: NigE, HKE, JamE, BrE, and AmE in

GloWbE, namely the verb REMEMBER. Due to the impossibility of analyzing all the examples retrieved for this verb, I searched for REMEMBER + *to*-infinitive, REMEMBER + *-ing*, and REMEMBER + *that*-clause (syntactic queries: *remember\*\_v to*, *remember\*\_v \*\_ing\_v*, and *remember\*\_v that\_cs\** respectively). The resulting data show that in NigE the use of non-finite complement clauses (36.6%) is again lower than in the standard varieties (47.4% in AmE and 49.4% in BrE), and also lower than in other varieties of English as a second language (HKE with 48.5%) and as a second dialect (JamE with 48.6%). In other words, while the distribution between finite and non-finite patterns with the verb REMEMBER is similar in both native and non-native varieties of English, with values around 48%, NigE shows a less frequent use of non-finite patterns, as is also the case with the verb REGRET.

Together with the influence of French on NigE, given the low frequency of use of non-finite patterns in all the L2s and ESD analyzed (as compared to native varieties), I will now examine the potential role of transfer from substrate languages.

In Nigeria more than 500 languages are spoken, but I will concentrate exclusively on the three languages with the highest number of speakers (Hausa, Igbo, and Yoruba). As mentioned in section 2.3, Hausa only makes use of *that*-complements and infinitives (Newman, 2000: 97), Igbo does not have clausal complementation (complements are always nominalized; cf. Emenanjo, 1987: 130), and Yoruba has *that*-complements and only one non-finite marker (Sheehan and van der Wal, 2016: 352). Since the main substrate languages do not have gerund forms, and, in the case of Igbo, not even clausal complementation, the use of the gerund forms may pose difficulties for speakers of NigE. Hence, the influence of substrate languages may play a role in the low proportion of non-finite forms found here.

Regarding HKE, the substrate language (Cantonese) does not have a distinction between finite and non-finite forms since non-finite forms are not available (Matthews and Yip, 1994: 174, 293). This absence of non-finite forms in Cantonese may present difficulties for HKE speakers in acquiring the different patterns of complementation. As noted above, according to Green's cline of the development of the clause system (2017: 173), the first patterns that a speaker learns are the finite ones, non-finite ones typically being the last to be learnt. Given that this variety straddles phase 2 and 3, we might expect that the influence of the substrate language is still strong, and that it justifies in part the low use of non-finite complements.

Finally, Jamaican Creole (the substrate language of JamE) does not have gerund forms or markers (Patrick, 2004: 423-424), so we would expect to have a distribution between finite and non-finite patterns similar to the other ESL varieties studied. However, its distribution is in fact very similar to that of native varieties (65.4% of non-finite in JamE vs. 68.9% of non-finite in AmE/BrE), which argues against the hypothesis of transfer processes taking place in L2 varieties of English.

As mentioned in section 2.3, the fact that the three non-native varieties under study are contact varieties means that variation might also be conditioned by contact-induced phenomena. One prototypical characteristic of languages learnt in contact situations is increased isomorphism (Schneider, 2012: 66; Green, 2017: 169) and hyperclarity

(Williams, 1987: 178). Finite patterns explicitly encode the relationship to the main clause through the use of the complementizer *that* when this is present, but also through tense, aspect and modality (Givón, 1985: 200). Therefore, finite patterns are more transparent and isomorphic than non-finite ones. According to Green's first language acquisition clause development (2017: 173), these finite patterns are also easier to acquire than the non-finite ones, as shown by the fact that they are acquired at an earlier stage of the learning process. The fact that JamE exhibits a higher proportion of non-finite forms, and is also the most advanced variety in Schneider's Dynamic Model (2007), lends support to the idea that it may have reached a state in which non-finite complementation is no longer obscure and complex for speakers, so that the increased isomorphism typical of language-contact situations is not so conspicuous as in NigE and HKE. As L2 varieties evolve, as in the case of JamE, speakers overcome the difficulties that non-finite patterns entail, and thus their use increases; hyperclarity is no longer needed.

Within the finite patterns, the distribution of *that*-complement clauses and zero-complement clauses can also be indicative of increased explicitness and isomorphism. In his study of different high frequency verbs, Schneider (2012: 83) finds that "the complementizer *that* is mostly more frequent in the New Englishes [WEs] than in GB [BrE]". In his data, the non-native or ESL varieties tend to be simpler and more isomorphic than the native variety that he considered. Rohdenburg (1996: 160) also studies *that* vs. zero finite patterns in native varieties of English and shows a correlation between higher complexity in the sequence and use of *that*. For example, the presence of negative markers in the complement and the presence of intervening material between the main clause and the complement clause add complexity to the processes of codification and decodification of the utterance. Examples in (29) exemplify this increase in complexity by the placement of an adverbial element between the two clauses, which is likely to trigger the use of the complementizer *that*, as can be seen in example (29a).

- (29) a. He told me (yesterday) that John had gone away.  
 b. He told me (yesterday) John had gone away. (Rohdenburg, 1996: 160)

It seems reasonable, then, to hypothesize that the contact-language situation may increase the complexity for the speaker and the hearer of English as an L2. In order to make themselves as clear as possible, speakers of L2 varieties of English would favor the use of the *that*-complementizer.

To confirm this, I examined the distribution of *that*- and zero-complement clauses in the varieties under study. As can be seen in Table 6, my data are in agreement with Schneider's (2012) claim: the presence of the complementizer ranges from 86% to 97% in L2 varieties, whereas in the native varieties the complementizer *that* is only present in 81.5% of cases. With this particular variant, *that* vs. zero, it is HKE that exhibits the highest frequency of the use of *that* (96.9%), that is, the variety that requires higher explicitness, since it is also the variety at the earliest stage of development of all the L2s studied here. NigE (86.5% of *that*), which showed a clear preference for finite (explicit)

patterns, and JamE (87.5% of *that*), which did not show such a strong preference for finite patterns, do not favor the use of explicit *that* as strongly as HKE. This might lead us to think that the preference for the use of the complementizer *that* in these two varieties may not have to do with the need for explicitness observed in HKE. In fact, this need for explicitness is not found in JamE with regard to the use of (explicit) finite patterns, since these are less frequent in JamE than in HKE and also NigE. Other factors such as influence of substrate languages and other transfer phenomena might be involved.

	AmE/BrE		JamE		HKE		NigE	
	n.	%	n.	%	n.	%	n.	%
<i>That</i>	317	81.5	49	87.5	63	96.9	154	86.5
<i>Zero</i>	72	18.5	7	12.5	5	7.4	24	13.5
<b>TOTAL</b>	<b>389</b>	<b>100</b>	<b>56</b>	<b>100</b>	<b>68</b>	<b>100</b>	<b>178</b>	<b>100</b>

Table 6: Distribution of *that*- and zero-complement clauses

A chi-square test was performed to determine whether there was a significant difference between the four groups of speakers with regard to the use or omission of the complementizer *that*. The chi-square statistic was not significant at the  $p < .05$ , p-value was .07138547.

This section has accounted for all the attestations of the verb REGRET retrieved in each variety, with special emphasis on the factors that determine the variation found, namely influence from substrate languages and general principles of transparency and increased isomorphism.

## 5. Conclusions

In this paper, I have studied and compared the clausal complementation systems of two supranational varieties (AmE and BrE), two ESL varieties (HKE and NigE) and one ESD variety (JamE). In an analysis of data from the GloWbE corpus, I considered all the influencing factors that might be at play in a given language-contact situation and that might serve to account for the differences found in the complementation of the verb REGRET in these WEs, namely historical and sociolinguistic factors, encapsulated in Schneider's Dynamic Model (2007), the influence of the substrate languages, and cognitive factors arising from contact situations.

One notable initial finding is the use of a prepositional phrase as a complement of the verb REGRETAS in *you will not regret about your employment* and *you will not regret for dating with military personnel*. This use of prepositional phrases to complement the verb REGRETIS not mentioned in FrameNet or in English grammars, and may be triggered by analogy with other structures, such as the noun REGRETUSED with prepositional phrases (*regret for the loss of a servant*) or the prepositional gerunds available with other types of verbs (*she delighted in doing it*). The tendency of the

prepositions *about* and *for* to co-occur with the verb REGRET, witnessed also in all the varieties under study here, may also be a sign of this verb becoming a prepositional verb. All these hypotheses, however, remain to be confirmed in future research.

As we have seen, native varieties of English make greater use of non-finite patterns than non-native varieties. Of the three non-native varieties studied here, the JamE complementation system comes closest to that of the supranational varieties, with a distribution of finite and non-finite patterns very similar to those of AmE and BrE. On the contrary, in HKE, and especially in NigE, this distribution shows a far lower proportion of use of non-finite patterns, reaching almost a 50-50 distribution in NigE.

Such differences correlate, on the one hand, with the evolutionary phases of Schneider's Dynamic Model (2007) in which the non-native varieties find themselves: JamE is currently in endonormative stabilization (phase 4) and HKE and NigE are at the nativization stage (both in phase 3, and HKE with traces of phase 2). Therefore, historical and sociolinguistic factors may be important determinants of the variation found.

On the other hand, I have also discussed evidence for the influence of transfer from the substrate languages. Firstly, in Cantonese, the substrate of HKE, subordination is constructed through parataxis and there is no distinction between finite and non-finite forms. The absence of the non-finite patterns could be partially transferred to HKE, thus explaining the low frequency of these patterns found in the data. Secondly, NigE may be influenced by French learnt at school as a foreign language and spoken in surrounding countries (Benin, Niger, Chad, and Cameroon), and by native Nigerian languages, the three most spoken of these being Hausa, Igbo and Yoruba. Both French and Hausa express complementation by using *that*-complements or infinitives, whereas Igbo's complements are always nominalized, and Yoruba only has *that*-clauses and one non-finite marker. In brief, none of these four languages has gerund forms, which could explain the low frequency of use of this form in NigE. Thirdly, Jamaican Creole has both *that*-clauses and infinitives as complement types. However, there are no gerund forms, meaning that transfer processes from Jamaican Creole to JamE in terms of the complementation system are also possible. Since Jamaican Creole does not have gerund, a low frequency of use of gerunds in JamE would be expected. However, this is not the case, and the use of non-finite forms in JamE is relatively high, probably because JamE is in an advanced phase of evolution which has overcome the difficulty that non-finite complement clauses pose.

In sum, substrate languages from the two ESL varieties (HKE and NigE) and the ESD variety (JamE) could have an influence on the clausal complementation system of the Englishes spoken in each region. The fact that the substrates do not have gerund forms may be partially transferred to English, with a consequent impact on the frequency of use of gerund forms. However, the fact that the JamE complementation system is very similar to that of the native varieties, even though Jamaican Creole does not have gerund forms, seems to invalidate the hypothesis of transfer processes. Hence, the data here seem to show that sociohistorical factors override substrate influences in advanced L2s. Of all the L2 varieties considered here, the most evolved variety in terms

of Schneider's Dynamic Model (2007), JamE, is the one with a highest frequency of use of non-finite patterns.

Finally, the influence of cognitive processes that take place in language-contact situations and SLA, such as hyperclarity (Williams, 1987: 178) and isomorphism (Schneider, 2012: 66; Green, 2017), seems to be another important factor determining the complementation systems of L2 varieties of English. On the one hand, finite patterns explicitly encode their relationship with the main clause through the use of the complementizer and the marked tense, aspect and modality (Givón, 1985: 200), which makes them more transparent and isomorphic. On the other hand, as Green (2017) demonstrates in her cline of the clause hierarchy, they are easier to acquire and thus are acquired earlier. As the non-native varieties evolve, according to Schneider's Dynamic Model (2007), non-finite complementation is no longer obscure and difficult for speakers and the use of this pattern in a wider range of situations increases. For example, JamE, being the most evolved non-native variety here, exhibits a higher proportion of non-finite patterns; however, HKE and NigE, being at an earlier stage in Schneider's model, exhibit a clear preference for more explicit, transparent, and isomorphic patterns, that is, finite patterns.

The distribution of the finite patterns (*that* and zero) also supports the hypothesis that cognitive factors are at work, since *that*-complement clauses are more explicit and isomorphic than their zero-complement counterpart. Even though the frequency of use of zero-complement clauses is low in all the varieties under study, its use is lower in the non-native varieties. There is a clear preference for the use of the *that*-complementizer in L2 varieties, reaching almost 97% in HKE. This may be due to the tendency for hyperclarity that these varieties display. There are even examples of *that* + to-infinitive and *that* + *-ing* in HKE, which would illustrate the tension between explicitness and economy (Slobin, 1983: 249); that is, the user knows how to form non-finite patterns, yet the tendency for hyperclarity and isomorphism leads them to use the complementizer anyway. However, it could also be analyzed as an adjectival or participial use of REGRET. In the case that more examples like those are found in future research all these possible analyses will have to be considered.

Through these findings I hope to have shown that verbal clausal complementation is an interesting area for the study of the interplay of factors that determine variation in different varieties of English.

## Notes

1. Research for this paper was funded by the Spanish Ministry of Economy and Competitiveness (grant FFI2014-53930-P) and the Regional Government of Galicia (grant ED431C 2017-50). I am also grateful to Elena Seoane for her insightful comments and encouragement, and to two anonymous reviewers as well as the editors for their helpful criticisms and suggestions.

2. FrameNet is a lexical database of English which annotates examples according to their use in actual texts and provides their meaning, usage and valence information (FrameNet).

3. I am aware that the status of this NP is debatable. Quirk et al. (1985: 1194), for example, consider it to be the subject of the non-finite clause, though they also refer to it as a raised object (1985: 1202). At the same time, Huddleston and Pullum et al. (2002: 1189-1190) consider both the genitive and the accusative as subject of the *-ing*-clause. Proof of the subject status of the NP preceding the *-ing* form is the fact that it becomes the subject of the corresponding finite clause, cf. *George regretted more than ever his forgetting of Nigel* → *George regretted more than ever that he forgot Nigel*.

4. Even though they express the same meaning the difference between preposition + NP and preposition + *-ing* is relevant, as one reviewer rightly pointed out, and it will be tackled in future research.

## Sources

Cambridge Dictionary. <<http://dictionary.cambridge.org/>>. Accessed 18 Apr. 2017.

FrameNet. “FrameNet Project”. <<https://framenet.icsi.berkeley.edu/>>. Accessed 3 May 2017.

GloWbE. Davies, Mark (2013): Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries. <<http://corpus.byu.edu/glowbe/>>. Accessed 8 Feb. 2017.

ICE. *International Corpus of English*. <<http://ice-corpora.net/ICE/INDEX.HTM>>. Accessed 30 July 2015.

Linguee. <<http://www.linguee.es/>>. Accessed 19 July 2017.

Merriam-Webster Dictionary. <<https://www.merriam-webster.com/>>. Accessed 18 Apr. 2017.

Oxford Dictionaries Online. <<http://www.oxforddictionaries.com/>>. Accessed 18 Apr. 2017.

## References

Allsopp, Richard, ed. (1996): *Dictionary of Caribbean English Usage*. Oxford: Oxford University Press.

Bernaisch, Tobias (2013): “The verb-complementational profile of *offer* in Sri Lankan English”. In Magnus Huber and Joybrato Mukherjee, eds., *Corpus Linguistics and Variation in English: Focus on Non-Native Englishes*: <http://www.helsinki.fi/varieng/series/volumes/13/bernaisch/> (17 October 2014). Helsinki: Research Unit for Variation, Contacts and Change in English.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999): *Longman Grammar of Spoken and Written English*. London: Longman.

Cuyckens, Hubert, Frauke D'hoedt and Benedikt Szmrecsanyi (2014): “Variability in verb complementation in Late Modern English: Finite vs. non-finite patterns”. In Marianne Hundt, ed., *Late Modern English Syntax*. Cambridge: Cambridge University Press, 182-204.

Davies, Mark (2012): “Examining recent changes in English: Some methodological issues”. In Terttu Nevalainen and Elizabeth C., eds., *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 263-287.

Davies, Mark and Robert Fuchs (2015): “Expanding horizons in the study of World Englishes with the 1.9 billion words Global Web-based English Corpus (GloWbE)”. *English World-Wide*, 36(1): 1-28.

- De Smet, Hendrik (2008): "Functional motivations in the development of nominal and verbal gerunds in Middle and Early Modern English". *English Language and Linguistics*, 12: 55-102.
- De Smet, Hendrik (2009): "Analysing reanalysis". *Lingua*, 119: 1728-1755.
- De Smet, Hendrik (2010): "English *ing*-clauses and their problems: The structure of grammatical categories". *Linguistics*, 48: 1153-1193.
- De Smet, Hendrik (2013): *Spreading Patterns: Diffusional Change in the English System of Complementation*. Oxford: Oxford University Press.
- De Smet, Hendrik (2014): "Constrained confusion: The gerund/participle distinction in Late Modern English". In Marianne Hundt, ed., *Late Modern English Syntax*. Cambridge: Cambridge University Press, 224-238.
- Denison, David (2017): "Counterfactuals". Paper Delivered at Santiago-Leuven-Edinburg Seminar Grammatical Variation and Change in English. KU Leuven, 3 April.
- Deshors, Sandra C. (2014): "Towards an identification of prototypical non-native modal constructions in EFL: A corpus-based approach". *Corpus Linguistics and Linguistic Theory*, 11(1): 19-50.
- 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.
- 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.
- Emenanjo, E. Nolue (1987): *Elements of Modern Igbo grammar: A Descriptive Approach*. Oxford: Oxford University Press.
- Fanego, Teresa (1990): "Finite complement clauses in Shakespeare's English, Part 2". *Studia Neophilologica*, 62: 129-149.
- Fanego, Teresa (1992): *Infinitive Complements in Shakespeare's English*. Santiago de Compostela: Servizo de Publicacións da Universidade.
- Fanego, Teresa (1996a): "The development of gerunds as objects of subject-control verbs in English (1400-1760)". *Diachronica*, 13: 29-62.
- Fanego, Teresa (1996b): "The gerund in Early Modern English: Evidence from the Helsinki Corpus". *Folia Linguistica Historica*, 17: 97-152.
- Fanego, Teresa (1996c): "On the historical development of the English retrospective verbs". *Neophilologische Mitteilungen*, 97: 71-79.
- Fanego, Teresa (1998): "Developments in argument linking in Early Modern English gerund phrases". *English Language and Linguistics*, 2: 87-119.
- Fanego, Teresa (2004a): "On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds". *Diachronica*, 21: 5-55.
- Fanego, Teresa (2004b): "Some strategies for coding sentential subjects in English: From exaptation to grammaticalization". *Studies in Language*, 28: 321-361.
- Fanego, Teresa (2007): "Drift and development of sentential complements in British and American English from 1700 to the Present Day". In Javier Pérez-Guerra, 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 Raymond Hickey, ed., *Eighteenth-century English*. Cambridge: Cambridge University Press, 200-220.
- Fanego, Teresa (2016): "The Great Complement Shift revisited: The constructionalization of ACC-ing gerundives". *Functions of Language*, 23(1): 84-119.
- Federal Ministry of Education, Nigeria (2004): *National Policy on Education*, 4<sup>th</sup>edn. Lagos: NERDC Press.
- Fischer, Olga (1988): "The rise of the for NP to V construction: An explanation". In Graham Nixon and John Honey, eds., *A Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*. London: Routledge, 67-88.
- Fischer, Olga (1989): "The origin and spread of the accusative and infinitive construction in English". *Folia Linguistica Historica*, 8: 143-217.
- Givón, Thomas (1985): "Iconicity, isomorphism and non-arbitrary coding in syntax". In John Haiman, ed., *Iconicity in Syntax*. Amsterdam/Philadelphia: John Benjamins, 187-219.
- Green, Clarence (2017): *Patterns and Development in the English Clause System: A Corpus-based Grammatical Overview*. Singapore: Springer.
- Hansen, Maj-Britt Mosegaard (2016): *The Structure of Modern Standard French: A Student Grammar*. Oxford: Oxford University Press.
- Huddleston, Rodney, Geoffrey K. Pullum et al. (2002): *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hundt, Marianne (2016): "Error, feature, (incipient) change – or something else altogether? On the role of low-frequency deviant patterns for the description of Englishes". In Elena Seoane and Cristina Suárez-Gómez, eds., *World Englishes: New Theoretical and Methodological Considerations* (Varieties of English Around the World G57). Amsterdam: John Benjamins, 37-60.
- Kachru, Braj B. (1985): "Standards, codification and sociolinguistic realism: The English language in the outer circle". In Randolph Quirk and Henry Widdowson, eds., *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press, 11-30.
- Loureiro-Porto, Lucía (Forthcoming): ICE vs GloWbE: A critical approach to big data as an alternative source for corpus compilation. *World Englishes*.
- Los, Bettelou (2005): *The Rise of the to-infinitive*. Oxford: Oxford University Press.
- Mair, Christian (2006): "Nonfinite complement clauses in the nineteenth century: The case of remember". In Merja Kytö, Mats Rydénand and Erik Smitterberg, eds., *Nineteenth-century English: Stability and Change*. Cambridge: Cambridge University Press, 215-228.
- Mair, Christian (2015): "Responses to Davies and Fuchs". *English World-Wide*, 36(1): 29-33.
- Matthews, Stephen and Virginia Yip (1994): *Cantonese: A Comprehensive Grammar*. London/New York: Routledge.
- Miller, D. Gary (2002): *Nonfinite Structures in Theory and Change*. Oxford: Oxford University Press.
- Mukherjee, Joybrato and Sebastian Hoffman (2006): "Describing verb-complementation profiles of New Englishes: A pilot study of Indian English". *English World-Wide*, 27(2): 147-173.
- Mukherjee, Joybrato and Stephan Th. Gries (2009): "Collostructional nativisation in New Englishes. Verb-construction associations in the International Corpus of English". *English World-Wide*, 30(1): 27-51.

- Mukherjee, Joybrato and Marco Schilk (2008): "Verb-complementation profiles across varieties of English". In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen, eds., *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: John Benjamins, 163-181.
- Mukherjee, Joybrato and Marco Schilk (2012): "Exploring variation and change in New Englishes: Looking into the International Corpus of English (ICE) and beyond". In Terttu Nevalainen and Elizabeth C. Traugott, eds., *The Oxford Handbook of the History of English*. Oxford: Oxford University, 189-199.
- Newman, Paul (2000): *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven/London: Yale University Press.
- Ogunmodimu, Morakinyo (2015): "Language policy in Nigeria: Problems, prospects and perspectives". *International Journal of Humanities and Social Sciences*, 5(9): 154-160.
- Olavarria de Ersson, Eugenia and Phillip Shaw (2003): "Verb complementation patterns in Indian Standard English". *English World-Wide*, 24(2): 137-161.
- Ouhalla, Jamal and Ur Shlonsky (eds.) (2002): *Themes in Arabic and Hebrew Syntax*. Dordrecht, The Netherlands: Kluwer Academic Publishers
- Patrick, Peter L. (2004): "Jamaican Creole: Morphology and syntax". In Bernd Kortmann and Edgar W. Schneider, eds., *A Handbook of Varieties of English. Vol 2: Morphology and Syntax*. Berlin: De Gruyter Mouton, 407-438.
- 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*, 76: 367-388.
- Rohdenburg, Günter (1996): "Cognitive complexity and increased grammatical explicitness in English". *Cognitive Linguistics*, 7(2): 149-182.
- Rohdenburg, Günter (2006): "The role of functional constraints in the evolution of the English complementation system". In Christiane Dalton-Puffer, Nikolaus Ritt, Herbert Schendl and Dieter Kastovsky, eds., *Syntax, Style and Grammatical Norms: English from 1500–2000*. Frankfurt: Lang, 143-166.
- Rohdenburg, Günter (2014): "On the changing status of *that*-clauses". In Marianne Hundt, ed., *Late Modern English Syntax*. Cambridge: Cambridge University Press, 155-181.
- Rudanko, Juhani (1998): *Change and Continuity in the English Language: Studies on Complementation over the Past Three Hundred Years*. Lanham, MD: University Press of America.
- Rudanko, Juhani (2000): *Corpora and Complementation: Tracing Sentential Complementation Patterns of Nouns, Adjectives and Verbs over the Last Three Centuries*. Lanham, MD: University Press of America.
- Rudanko, Juhani (2011): *Changes in Complementation in British and American English: Corpus-based Studies on Non-finite Complements in Recent English*. Basingstoke: Palgrave.
- Sapir, Edward (1921): *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Co.
- Schneider, Edgar W. (2003): "The dynamics of New Englishes: From identity construction to dialect birth". *Language*, 79(2): 233-281.
- Schneider, Edgar W. (2007): *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

- Schneider, Edgar W. (2012): "Contact-induced change in English worldwide". In Terttu Nevalainen and Elizabeth C. Traugott, eds., *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 572-581.
- Schneider, Edgar W. (2013): "English as a contact language: The 'New Englishes'". In Daniel Schreier and Marianne Hundt, eds., *English as a Contact Language* (Studies in English Language). Cambridge: Cambridge University Press, 131-148.
- Seoane, Elena (2016): "World Englishes today". In Elena Seoane and Cristina Suárez-Gómez, eds., *World Englishes: New Theoretical and Methodological Considerations* (Varieties of English around the World G57). Amsterdam/Philadelphia: John Benjamins, 1-15.
- Sheehan, Michelle and Jenneke van der Wal (2016): "Do we need abstract case?" In Kyeongmin Kim, Pocholo Umbal, Trevor Block, Queenie Chan, Tanie Cheng, Kelli Finney, Mara Katz, Sophie Nickel-Thompson and Lisa Shorten, eds., *Proceedings of the 33rd West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 351-360.
- Slobin, Dan (1980): "The repeated path between transparency and opacity in language". In Ursula Bellugi and M. Studdert-Kennedy, eds., *Signed and Spoken Language: Biological Constraints on Linguistic Form*. Weinheim: Verlag Chemie, 229-243.
- Slobin, Dan (1983): "What the natives have in mind". In Roger Andersen, ed., *Pidginization and Creolization as Language Acquisition*. Rowley, MA: Newbury House, 246-253.
- Steger, Maria and Edgar Schneider (2012): "Complexity as a function of iconicity: The case of complement clause constructions in New Englishes". In Bernd Kortmann and Benedikt Szmrecsanyi, eds., *Linguistic Complexity: Second Language Acquisition, Indeginization, Contact*. Berlin: De Gruyter Mouton, 156-191.
- Strang, Barbara M. H. (1970): *A History of English*. London: Methuen.
- Thomason, Sarah G. (2001): *Language Contact: An Introduction*. Washington, D.C.: Georgetown University Press.
- Thomason, Sarah G. (2008): "Social and linguistic factors as predictors of contact-induced change". *Journal of Language Contact*, 2: 42-56.
- Vosberg, Uwe (2006): *Die Große Komplementverschiebung: Außersemantische Einflüsse auf die Entwicklung satzwertiger Ergänzungen im Neuenglischen*. Tübingen: Narr.
- Warner, Anthony (1982): *Complementation in Middle English and the Methodology of Historical Syntax: A Study of the Wyclifite Sermons*. London: Croom Helm.
- Williams, Jessica (1987): "Non-native varieties of English: A special case of language acquisition". *English World-Wide*, 8(2): 161-199.