

# A dialectological approach to complement variability in global web-based English

Raquel P. Romasanta  
University of Santiago de Compostela / Spain

**Abstract** – Computer-Mediated Communication is part of the everyday lives of a great many people of all ages, cultures, social statuses, and geographical locations. In the present study, I explore non-categorical syntactic variability in internet language with data from the *Corpus of Global Web-Based English* (GloWbE), which includes material from blogs, forums, comments, and other types of websites. The focus is on how the geographical area of internet users affects the use of the clausal complementation patterns available for the verb REGRET. The analysis of more than 10,000 examples from Indian, Sri Lankan, Pakistani, Bangladeshi, Singaporean, Malaysian, Philippine, Hong Kong, British, and American Englishes shows that geographical origin does have a bearing on the complementation system of this verb, in terms of both the factors that determine variability and the preferences for particular patterns. The varieties displaying more similarities are those that are geographically close, making the distinction between three geographical areas possible: South Asia (India, Sri Lanka, Pakistan, and Bangladesh), South-East Asia (with Singapore, Malaysia, and the Philippines) and East Asia (Hong Kong).

**Keywords** – computer-mediated communication; complementation; World Englishes; language contact; geographical proximity; transfer.

## 1. INTRODUCTION<sup>1</sup>

Santoro (1995: 11) defines Computer-Mediated Communication (henceforth, CMC) as encompassing all computer uses, including statistical and financial programs, remote-sensing systems, and so on, and Herring (1996: 1) defines it as “communication that takes place between human beings via the instrumentality of computers.” Nowadays, when we talk about CMC, we focus mainly on the communication through and about the internet and web, including instant messaging, video conference, email, social media, and the World Wide Web. This work draws on data from the web for the study of a grammatical construction across Englishes around the world, in particular, clausal complementation

---

<sup>1</sup> I would like to express my appreciation to the two anonymous reviewers and the editors whose constructive comments improved the quality of the paper considerably. Any errors remain my sole responsibility. For support with this study, my gratitude goes to the *Spanish Ministry of Science and Innovation* (grant PID2020–117030GB–I00 funded by MCIN/AEI/10.13039/501100011033), and the *Recovery, Transformation, and Resilience Plan of the European Union NextGenerationEU* (University of Vigo, grant ref. 585507).



after the verb REGRET. Data is taken from the *Corpus of Global Web-Based English* (GloWbE; Davies and Fuchs 2015a), which includes blogs, forums, comments, and other types of websites from 20 different countries.

A previous study (Romasanta 2021) in Asian varieties on the complementation profile of REGRET, which allows non-categorial variation between finite (*that*) and nonfinite (*-ing*) complement patterns with anterior (1) and simultaneous (2) meanings, finds similar distributions of complements across varieties.

- (1) a. This is when you will hugely **regret** *that* you went to Lahore to attend your second cousin's...
  - b. This is when you will hugely **regret** *going* to Lahore to attend your second cousins third marriage to a half Iranian-half Pakistani woman brought up in the US, because you thought it would be a lark. (GloWbE-BD)
  
- (2) a. In these circumstances the Secretary of State **regrets** *that* he is not prepared to extend your stay to enable you to continue as a student at one of the Hubbard Colleges. (GloWbE-HK)
  - b. In these circumstances the Secretary of State **regrets** not *being* prepared to extend your stay to enable you to continue as a student at one of the Hubbard Colleges.

For example, Pakistani and Sri Lankan Englishes have a clear preference for finite patterns, with 57 percent and 55 percent, respectively, and Hong Kong, Bangladeshi, and Indian Englishes prefer nonfinite complements, with 59 percent, 59 percent, and 61 percent, respectively. The author hypothesizes that these similarities might be explained by the complement constructions available in the substrate languages spoken in each region since many times the effects of language contact do not surface as direct structural transfer from the indigenous languages to the target language, but rather as differences in frequencies of use and preference for some patterns over others, which makes its identification more difficult (see also Thomason 2001; Gut 2011; Brunner 2014, 2017; Romasanta 2021). Romasanta (2021: 1162) concludes that the substrate languages do not seem to explain the similarity of distributions since varieties with the same complementation systems show similar distributions to those with different systems. Other hypotheses briefly mentioned in the study without any statistical tests applied are the evolutionary development of the individual varieties and geographical proximity.

The present study focuses on the latter hypothesis, geographical proximity of English varieties. That is, on how the geographical area of internet users affects the use

of the clausal complementation patterns available for this verb, not only in terms of the distribution of the patterns but also the intra-linguistic conditioning factors affecting the speakers' choice. Regarding the geographical areas, I distinguish between South Asia (India, Sri Lanka, Pakistan, and Bangladesh), South-East Asia (Singapore, Malaysia, and the Philippines), and East Asia (Hong Kong), and the United States and Great Britain as a baseline. The aim is to test the fundamental principle of dialectology that states that “geographical proximity between dialects should predict dialectal similarity between dialects” (Szmrecsanyi 2013: 837). In other words, we can expect geographically close varieties to exhibit more similarities than distant ones, and, in principle, this should be the case for the language used on the internet, as it is elsewhere. Therefore, South Asian varieties should exhibit more similar complementation preferences when compared to the South-East Asian ones. A study of the distribution of the aforementioned finite and nonfinite complementation patterns, not only in general numbers but also in terms of the factors that influence the choice through non-hierarchical phylogenetic networks (*NeighborNet*; Bryan and Moulton 2004), will help me to test this principle.

The paper is structured as follows. Section 2 provides an overview of the extra-linguistic factors that might be at play in syntactic variability, i.e., geographical proximity, second language acquisition (henceforth SLA) processes (transparency and transfer from substrate languages), and evolutionary phase of development. Section 3 describes the data selection, annotation, and analysis. Section 4 discusses the results of the study and is followed by the conclusion in Section 5.

## 2. THEORETICAL BACKGROUND

English varieties around the world, or World Englishes, have been described in the literature as independent varieties of English in their own right, as opposed to simple deviations from British English (Platt *et al.* 1984), and as exhibiting similarities to other English varieties (Stevens 1980: 85).

The study of geography as a determining factor of similarities across dialects, although frequently neglected in the study of World Englishes, is a common practice in dialectology studies and one of the extra-linguistic dimensions along which English varieties are commonly aligned (Szmrecsanyi and Röthlisberger 2019; Szmrecsanyi and Grafmiller 2023). The focus of the present study is not the analysis of dialects of English

in the traditional sense; however, it seems plausible that geographical proximity might also predict similarity between varieties of English. In studies on World Englishes, this was raised as early as 1980 in Strevens's *World Map of English Model*, in which he mentions that each form of English "normally exhibits similarities with other forms of English in the same geographical area" (Strevens 1980: 85). However, there has been little work that considers geography as a potential predicting factor for similarities and dissimilarities across global varieties of English. Of the very few authors who have done so, Szmrecsanyi and Kortmann (2009b) and Szmrecsanyi (2013) find geography to be a weak predictor of variability. Szmrecsanyi (2013: 841), for example, in a study of morphosyntactic similarities in L1 varieties, finds that geography accounts for less than five percent of the variability found and that there is a typological split "between traditional L1 varieties, high-contact L1 varieties, and what we have dubbed 'higher-contact' L1 varieties of English (such as the AAVE varieties)." In contrast, Kortmann and Schröter's (2017: 308) *NeighborNet* analysis of the survey data from the *World Atlas of Variation in English* project yields evidence of regional clustering, for example, South Asian and South-East Asian varieties in the same cluster but in different branches. In this direction, Fuchs *et al.* (2019) look at the present perfect in African English varieties, British, American, and Philippine English, and also find geographical proximity as the most important predictor.<sup>2</sup>

In the remainder of this section, I will briefly describe the other extra-linguistic factors that might affect the English varieties around the world previously mentioned.

### 2.1. SLA and language contact processes

There are two main processes that I would like to discuss here: 1) the principle of maximization of transparency and 2) language transfer. The principle of maximization of transparency is one of the production principles mentioned by Williams (1987).<sup>3</sup> Slobin (1980) considers transparency as the one-to-one mapping of form and meaning, that is, an intended underlying meaning is expressed with one clear, "invariant surface form (or

---

<sup>2</sup> As a reviewer rightly pointed out, the highly active work on epicenter theory in World Englishes relates to this argument. However, as the present study focuses on language-use data, it will not be possible to identify the influence of a variety on another. In order to do so, a mixed-method approach, including attitudinal data as well as historical background data, is necessary (Hundt 2013: 184; Peters and Bernaisch 2022).

<sup>3</sup> Also referred to as the 'one-to-one principle' in Andersen (1984), 'iconicity' in Haiman (1985), and 'isomorphism' in Givón (1985).

construction)” (Andersen 1984: 79). World Englishes are said to show a tendency towards transparency because transparent constructions are easier for the speaker to produce and for the listener to parse (Slobin 1973, 1977; Karmiloff-Smith 1979; Williams 1987: 179). Multiple studies have focused on this tendency for transparency (see, for example, Williams 1987; Szmrecsanyi and Kortmann 2009a; Steger 2012; Romasanta 2017). In the complementation system, this was attested within the alternation between finite and nonfinite clauses. Finite complement clauses are more transparent because they are marked for tense, agreement and modality, have an explicit subject, and usually a complementizer, and therefore the relationship between form and meaning is tighter than in nonfinite clauses (Givón 1985: 200; Schneider 2012a, 2013; Steger and Schneider 2012; Romasanta 2017, 2019). Therefore, in the present study, I will test this tendency for transparency by looking at the distribution between finite and nonfinite patterns with the verb REGRET.

The other SLA and language contact process —and probably the most obvious contact-induced change— is transfer. At the level of grammar, Schneider (2007: 83) argues that innovations occur mainly at the interface between lexis and grammar, a classic example being verb and adjective complementation, and indeed a series of studies have focused on the innovations present in the complementation system (see Mukherjee and Hoffmann 2006; Mufwene and Gries 2009; Deshors and Gries 2016; Gries and Bernaisch 2016, among others). In order to find this language transfer, we must know the complementation systems of the different substrate languages spoken in each region. Methodologically, this brings up some difficulties. Firstly, it is impossible to assign a particular substrate language to a particular speaker in the GloWbE data, and, secondly, the number of substrate languages in some countries goes beyond the hundreds, so I could only look at the ones with a written tradition. This means that conclusions for the effect of language transfer must be taken with care. In what follows, I will briefly describe the complementation systems of the main substrate languages in each region, although we must not forget that the sociolinguistic situations of these regions are more complex than can be described in detail here (see Table 1 in Section 2.2 for a summary of the substrate languages and the phases of development of each variety).

Based on the *World Factbook* (CIA 2024), the dominant substrate languages in **India** are Hindi, Bengali, Marathi, and Telugu. Even though many other languages are also part of the sociolinguistic landscape —for example, Tamil, Gujarati, Urdu, Kannada,

Malayalam, Punjabi, among others—I will focus on the first four since they are the most widely spoken languages. All four languages (Hindi, Bengali, Marathi, and Telugu) use finite clauses in their complementation system, and these are marked with the complementizers *ki*, *bôle*, *ki*, and *ani*, respectively. Three of these languages (Hindi, Marathi, and Telugu) have nonfinite complements, which consist of the suffixes *na*:- in Hindi, *-aTam* in Telugu, and *-āy*, *-ūn*, and *-lyā* in Marathi, added to the verb stem (see Koul 2008: 181–185 for Hindi, Krishnamurti and Gwynn 1985: 234, 363 for Telugu, Pandharipande 1997: 65–68, 444 for Marathi).

In **Sri Lanka**, the main substrate languages are Sinhala and Tamil (CIA 2024). In Sinhala, finite complement clauses can be constructed with the complementizer *kiəla*, and nonfinite complements with the complementizers *bawə* and *ekə* with a nonfinite verb (Wheeler *et al.* 2005: 173–174). In Tamil, finite complements take the complementizer *nuu* (Schiffman 1999: 152, 174), and nonfinites are constructed adding the suffixes *-a*, *-tu*, *-ntu*, *-ttu*, or *-i* to the verb stem (Lehmann 1993: 71–72).

According to the *World Factbook* (CIA 2024), the dominant substrate languages in **Pakistan**, are Punjabi, Pashto, and Sindhi. Finite complements are constructed with the marker *ki* in Punjabi, *tse* or *che* in Pashto, and *ta* in Sindhi. The suffixes *-Naa/naa* and *-an.u/in.u*, in Punjabi and Sindhi, respectively, are used for nonfinite complementation (see Bhatia 1993: 44, 50 for Punjabi, Tegey and Robson 1996: 199 for Pashto, and Yegorova 1971: 74–75 for Sindhi).

Bengali, also known as Bangla, is the dominant substrate language in **Bangladesh** (CIA 2024). As mentioned previously, Bengali has only finite complements.

In **Singapore**, the dominant substrates are Mandarin and other Chinese dialects (including Hokkien, Cantonese, Teochew, Hakka; CIA 2024). These languages use the juxtaposition of clauses, so neither finite nor nonfinite complementation is possible here (see, for example, Haspelmath *et al.* 2001: 979 for Mandarin, Fang 2010: 104 for Hokkien, and Matthews and Yip 1994: 174, 293 for Cantonese).

In **Malaysia**, the main substrate languages are Malay and a number of Chinese dialects. As mentioned previously, the Chinese dialects do not have finite or nonfinite complementation. In Malay, finite complement clauses take the complementizer *bahawa* (Omar and Subbiah 1989: 97) while nonfinite clauses do not exist (Nordhoff 2009: 276–279).

According to the *World Factbook* (CIA 2024), the dominant substrate language in the **Philippines** is Tagalog, where finite clauses are introduced by the linker *na/-ng* (Schachter and Otones 1972: 172).

Cantonese is the dominant substrate language in **Hong Kong**, 88.9 percent, together with Mandarin and other Chinese dialects (CIA 2024). As already stated, these Chinese dialects do not have finite or nonfinite complementation.

## 2.2. *Evolutionary phase of development in the Dynamic Model (Schneider 2007)*

The most widely discussed model of classification of World Englishes is the ‘Dynamic Model’ (Schneider 2007).<sup>4</sup> The main assumption here is that the different post-colonial Englishes undergo the same uniform process of identity reconstruction divided into five phases: foundation, exonormative stabilization, nativization, endonormative stabilization, and differentiation (Schneider 2007: 30–35). Various earlier studies found a correlation between phase of development in this model and degree of complexity. Research on verb complementation in particular shows mixed results regarding this correlation. Mukherjee and Gries (2009: 48–49) study ditransitive, monotransitive, and intransitive constructions in Hong Kong, Indian, and Singaporean English showing that the correlation holds true: “the more advanced a New English variety is in its evolution, the more dissimilar it is to British English at the level of collocations.” Schneider (2012b) looks at the alternation between finite and nonfinite clauses with several number of verbs, taking into account the presence or absence of the complementizer *that* and an explicit modal. His results also confirm the correlation in that they indicate that less advanced varieties, in this case Hong Kong English and East African English, have a stronger tendency to use simpler patterns than the more advanced ones, Singaporean and Indian Englishes. However, the correlation is not found in Deshors and Gries’ (2016) study of *-ing* and *to*-infinitive complement alternation in Singaporean, Hong Kong, and Malaysian Englishes. The most advanced variety, Singaporean English, is not dissimilar, but in fact the most similar to the native Englishes (British and American English). In a similar vein, García-Castro (2018) and Romasanta (2019, 2021) study complement variability with the retrospective verbs REMEMBER and REGRET, respectively, and also detect stronger preferences for

---

<sup>4</sup> Other models of classification frequently alluded to are also available. For example, Kachru (1982), Mair (2013) and, more recently, Buschfeld and Kautzsch (2017).

simpler finite clauses in less advanced varieties and for more complex nonfinite patterns in the more advanced varieties. The greater use of nonfinite patterns in more advanced varieties, therefore, makes them more similar to British English.

It seems suitable then to briefly consider the evolutionary phase of development in the Dynamic Model of each Asian variety included in the study to assess the potential effect on the alternation between finite and nonfinite complementation. Table 1 below summarizes this. Two important notes are in point. Firstly, Singapore is in phase 4 in the Dynamic Model, endonormative stabilization. However, it is said to have become a first/native language (L1), with many of its young speakers learning it as their first language, so that it is gradually developing from ESL to ENL (Gupta 1994; Lim and Foley 2004; Tan 2014; Lim 2017; Buschfeld 2020a, 2020b). Secondly, regarding Hong Kong, Schneider (2007: 133) claims that it has “reached stage 3 [but] with some traces of phase 2 still observable,” and Setter *et al.* (2010: 116) argue that “Hong Kong English will eventually be pushed more firmly towards Kachru’s Outer Circle, Schneider’s phase 4.” Until the handover of the territory to China in 1997, English was the medium of instruction in most schools, but a change in policy then ensued. There has since been a process of mainlandization by which the government has begun to favor the use of Cantonese as the medium of instruction, while reducing the number of schools allowed to use English.

	Variety	Complementation			Evolutionary phase
		Finite	Nonfinite	Summary	
<b>South Asia</b>	India	Yes	Yes	Both	3+
	Sri Lanka	Yes	Yes	Both	4
	Pakistan	Yes	Yes	Both	3+
	Bangladesh	Yes	No	Finite	2+
<b>South-East Asia</b>	Singapore	No	No	None	4
	Malaysia	Yes	No	Finite	3
	The Philippines	Yes	?	Finite	4
<b>East Asia</b>	Hong Kong	No	No	None	3

Table 1: Summary of the substrate languages and the phase of development of each Asian variety of English



### 3. DATA AND METHODOLOGY

#### 3.1. *The corpus*

The data has been taken from the GloWbE corpus (Davies and Fuchs 2015a), an online corpus released in 2015 with 1.9 billion words from 1.8 million web pages in 20 different countries (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).

In order to identify the countries of origin of each web page, they carried out the searches for each country separately relying on *Google's Advance Search*, which relies on country domains as well as on “the IP address for the web server, who links to that website, and who visits the website” (Davies and Fuchs 2015b: 4). This, however, has been criticized several times since country domains such as *.to* (Tonga) may retrieve websites from Tokyo, Toronto, or Timbuctoo, as well as websites such as [www.knowhow.to](http://www.knowhow.to) or [www.invitation.to](http://www.invitation.to). Even if the website is correctly cataloged, the writer may not be originally from the country (Nelson 2015: 39; Deshors and Bernaisch 2019). This also has an impact on the researchers' knowledge of the writer's backgrounds (age, gender, mother tongue, etc.), which is especially relevant for the present study as one of the hypotheses is related to the substrate languages of the writers. From a methodological perspective, the study of the substrate languages poses a problem, and, therefore, conclusions on this matter are to be taken with care.

Despite of the issues mentioned above, I see the GloWbE corpus as “a big and aggregative corpus” (Brezina and Meyerhoff 2014; Mukherjee 2015: 36) and expect that its size will statistically overcome its hindrances (Davies 2012; Nelson 2015: 39; Hundt 2020). In fact, studies based on GloWbE that replicate earlier studies carried out with smaller corpora obtain similar results (see, for example, Heller and Röthlisberger 2015).

#### 3.2. *Manual data pruning and coding*

For this study, data represents eight different English varieties from the Asian continent, namely Indian, Sri Lankan, Pakistani, Bangladeshi, Singaporean, Malaysian, Philippine, and Hong Kong Englishes, and the two main metropolitan varieties, British and American English in the GloWbE corpus (regret\*\_v\*). The total number of examples retrieved was 10,275.

After the manual pruning of the examples, I codified all relevant instances according to 11 intra-linguistic conditioning factors. The list is as follows:

1. Meaning of the verb in the MC (main clause) (dichotomous: regret1, regret2).
2. Meaning of the verb in the CC (complement clause) (dichotomous: action, state).
3. Animacy of the subject in the CC (dichotomous: animate, inanimate).
4. Type of subject in the MC (qualitative: pron1, pron2, pron3, NP, none).
5. Type of subject in the CC (dichotomous: complex noun phrase (CNP), other).
6. Voice of the CC (qualitative: active, passive, copular).
7. Polarity of the CC (dichotomous: positive, negative).
8. Complexity of the CC (quantitative: number of words).
9. Presence of intervening material (quantitative: number. of intervening words).
10. Subject coreferentiality (dichotomous: coreferential, non-coreferential).
11. Horror aequi (dichotomous: *yes*, no).

The first three of these are semantic factors. The two meanings of the verb in the MC are taken from Cuyckens *et al.* (2014: 188) where they define ‘regret1’ as “to feel sorry about something one has done and that one should have done differently or about a state of affairs one is involved in or responsible for and that one wishes was different”, as in (3), and ‘regret2’ as a “a more ‘polite’ use of REGRET where the speaker says that s/he is sorry or sad about a situation, usually one that s/he is not directly responsible for,” as in (4). For the meaning of the verb in the CC, the distinction between action and state was drawn from Quirk *et al.* (1985: 201), see examples (5) and (6), respectively. Lastly, for the animacy of the subject in the CC, I used a binary classification distinguishing between animate (7) and inanimate (8).

- (3) Tepco conference starting now: “We **regret** that we are causing concern to many residents of Japan.” (GloWbE-US)
- (4) We **regret** that our client was not provided with more time. (GloWbE-LK)
- (5) One thing I know is that I never **regret** *attending* this course. (GloWbE-MY)
- (6) He **regrets** not *having* the chance to tour the Philippines yet, things that made him feel... (GloWbE-PH)

- (7) We do **regret** *that the terrorists* were actually horrific acts and they were terrorist acts. (GloWbE-PK)
- (8) Bradley has since publicly stated he was humbled by the Morton case and **regrets** *his actions* opposing DNA testing in the case. (GloWbE-US)

The next seven factors (from four to ten above) are features relating to processing complexity. These are important for the alternation, since with complement clauses involving higher processing complexity, speakers generally prefer more grammatically explicit constructional variants ('Complexity Principle'; Rohdenburg 1996). The complement clause can be active, passive, or copular, as in (9), (10), and (11), respectively, and negative (12) or positive. Then, I coded the complexity of the complement clause (13), which contains a total of 87 words, and the presence of intervening material (14), which has six words as intervening material. In terms of subject coreferentiality between the main and complement clauses, these can be coreferential (15) or non-coreferential (16). Finally, the last factor exemplifies the generalization known as the 'Horror Aequi Principle', which holds that speakers tend to avoid (near-)identical and (near-)adjacent structures (Brugmann 1909; Rohdenburg 2003). This factor has two levels, 'yes' (17), when there is an environment where this principle might be at work, and 'no'.

- (9) I **regret** that I have *wasted* about 2 weeks on this site trying to reason and arrive at some kind of consensus which would move Sri Lanka forward, ... (GloWbE-LK)
- (10) ... that they might have no cause to **regret** *being denied* the option of any other. (GloWbE-GB)
- (11) Have you ever **regretted** *being* a monk? (GloWbE-MY)
- (12) The meeting **regretted** that India was *not* interested in the resumption of dialogue. (GloWbE-PK)
- (13) I **regret** *that the U.S. has suffered itself to be brought so low by the vultures and crooks who are operating the roulette wheels and faro tables in the Fed, that is now obliged to throw itself on the mercy of its legislators and charwomen, its clerks, and its poor pensioners and to take money out of our pockets to make good the defalcations of the International Bankers who were placed in control of the Treasury and given the monopoly of U.S. Currency by the misbegotten Fed.* (GloWbE-US)
- (14) I **regret**, *from a personal point of view*, being here. (GloWbE-GB)
- (15) *She* constantly **regrets** that *she* could not afford to send her daughters to school during the hard times, ... (GloWbE-BD)
- (16) ... although *we* **regret** *her* not coming to Asia-Pacific, so that she could address this... (GloWbE-LK)

(17) I am **regretting** *writing* it but can't stop because you deserve it. (GloWbE-IN)

### 3.3. Statistical analysis

Data was subjected to non-hierarchical phylogenetic networks (*NeighborNet*) as an exploratory method to visually represent which varieties are more similar and whether this could correspond to geographical proximity. This is a clustering method originating in bioinformatics (Bryant and Moulton 2004) and frequently used in historical, dialectological, and typological linguistics (McMahon and McMahon 2005; Cysouw 2007; McMahon *et al.* 2007; Szmrecsanyi and Wolk 2011). These networks allow for a more fine-grained analysis, as compared to other multidimensional aggregation analyses such as hierarchical cluster analysis, as they “produce an unrooted network representation (*NeighborNet*) that establishes, first of all, “geolinguistic signal[s]” (Szmrecsanyi 2013) in the data” (Werner 2014).

The analysis was conducted in *R* (R Core Team 2022) using the *NeighborNet* package (Ansari and Draghici 2019). These have been shown to be a great tool to graphically represent relationships of similarity and dissimilarity between multiple objects. Each object, here English varieties, represents its own cluster. They are compared pairwise within a distance matrix and the most similar ones are merged until all objects are merged into one tree. In order to create the distance matrix, I used the relative values of the individual factors and, to measure distances and similarities between varieties of English, I used the Euclidean distance, which in the case of the present dataset is fully proportional to the Manhattan distance. The Euclidean distance measure “is similar to our everyday idea of the distance between two objects”, where we would take the shorter direct route (see Figure 1 below; Levshina 2015: 306–307). The resulting networks are unrooted family trees so that the length of each branch is proportional to linguistic distances (Bryant and Moulton 2004; Szmrecsanyi 2013: 841). This means that proximity in the net indicates similarity in the complementation profile of REGRET in the varieties of English.

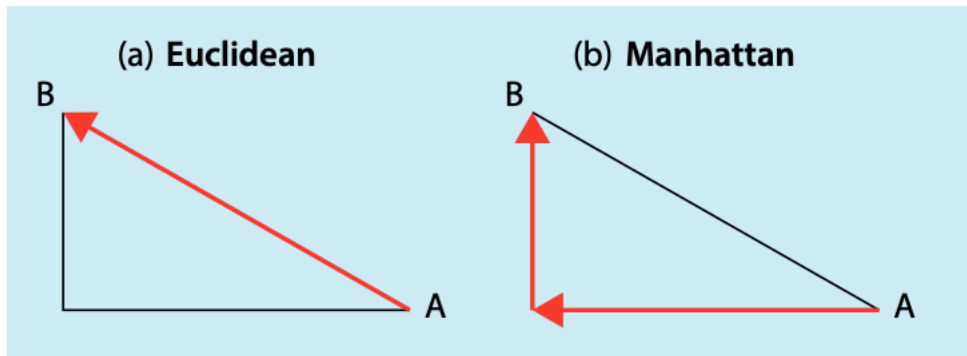


Figure 1: Distance metrics: a) Euclidean, b) Manhattan (from Levshina 2015: 307)

#### 4. RESULTS AND DISCUSSION

I begin the data analysis with an overview of the distribution of finite *that*-clauses and nonfinite *-ing* clauses across varieties. As can be seen in Figure 2, British (GB) and American English (US) have the same distribution, with a clear preference for nonfinite clauses (73%) over finite ones (27%). The next three varieties, Singaporean (SG), Malaysian (MY), and Philippine Englishes (PH) have a very similar distribution to the metropolitan varieties, or even the same as in the case of the Philippines. Compared to British and American Englishes, Singaporean English shows a slightly stronger preference for *-ing* clauses, 78 percent. Then, Malaysian and Philippine Englishes have a very similar distribution, with 74 percent and 73 percent of nonfinite clauses, respectively. The remaining varieties, Indian (IN), Bangladeshi (BD), Hong Kong (HK), Sri Lankan (LK), and Pakistani Englishes (PK) show a stronger use of *that*-clauses, with 39 percent, 41 percent, 41 percent, 54 percent, and 57 percent, respectively, as compared to British and American Englishes, which only have 27 percent of finite clauses.

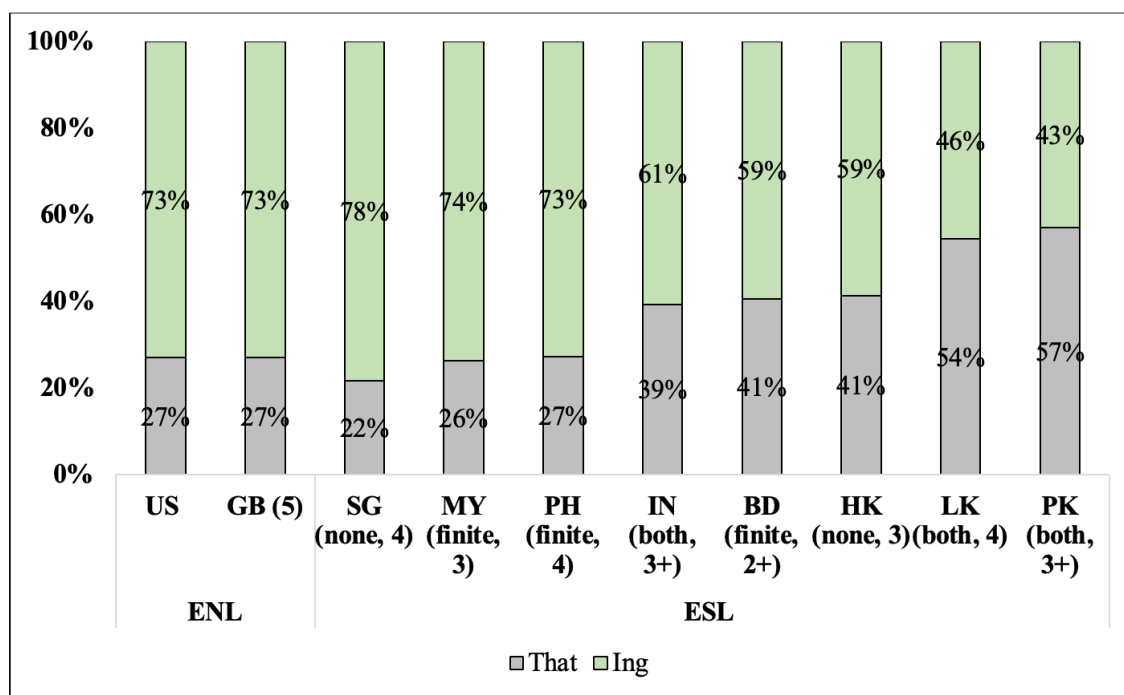


Figure 2: Distribution of finite *that*-clauses and nonfinite *-ing* clauses across Asian varieties of English

The greater use of finite patterns in India, Bangladesh, Hong Kong, Sri Lanka, and Pakistan might be the effect of the SLA strategy of maximization of transparency described in Section 2.1. According to this, ESL speakers would prefer transparent constructions due to these being easier to produce and parse. However, looking at the data in Figure 2, this tendency towards transparency would not explain why there is a stronger preference for *that*-clauses in some varieties. The explanation does not seem to lie on the transfer effect from substrate languages since, as can be seen in brackets, varieties with similar complementation systems in their substrate languages have different distributions between finite and nonfinite patterns. See, for example, the distributions in Singapore and Hong Kong. Complementation in the substrate languages in both regions is constructed through parataxis, that is, the juxtaposition of two clauses so that “the two clauses are more symmetrical than main and subordinate clauses in English” (Matthews and Yip 1994: 293). However, Singapore shows a clear preference for the use of *-ing* clauses (78%) while this preference is reduced to 59 percent in Hong Kong. The same occurs with India and Sri Lanka or Pakistan, where finite and nonfinite complement constructions are available in the substrate languages. While, in India, there is a stronger use of *-ing* clauses (61%), in Sri Lanka and Pakistan, the preference is for *that*-clauses (54% and 57%, respectively).

Another potential explanatory extra-linguistic factor mentioned in Section 2.2. is the effect of the evolutionary phase of development in terms of Schneider's Dynamic Model (2007). According to different studies, there should be a stronger preference for simpler *that*-clauses in less advanced varieties and for more complex *-ing* clauses in the more advanced varieties (Schneider 2012b; Brunner 2017; García-Castro 2018; Romasanta 2021). Looking back at Figure 2, less advanced varieties such as Hong Kong and Bangladesh, in phases 3 and 2+, have a stronger preference for simpler *that*-clauses (41%), as compared to Great Britain, with 27 percent. However, Malaysia, which is another variety in phase 3, shows a clear preference for more complex *-ing* clauses (74%). The more advanced varieties, Singapore and Philippines, both in phase 4, have a clear preference for the use of complex *-ing* clauses (78% and 73%, respectively), but other varieties, such as Sri Lanka and Pakistan, in phases 4 and 3+, prefer *that*-clauses, 54 percent and 57 percent, respectively. Therefore, the evolutionary phase of development does not seem to fully account for the different distributions of finite and nonfinite complement patterns across English varieties.

Figure 3 below is the output of the non-hierarchical phylogenetic network (*NeighborNet*) where distances between varieties are represented considering the distribution of finite and nonfinite patterns and the 11 intra-linguistic conditioning factors described in Section 3.1. Each node is one English variety, here referred to by their respective abbreviations, and information regarding the phase of development and the presence or absence of finite and/or nonfinite clauses in the dominant substrate languages in the parentheses. When there are finite and nonfinite complements in the substrates, I use 'both'. 'Finite' is used when only finite clauses are possible, and 'none' when complementation is constructed through other strategies and neither finite nor nonfinite complements exist. The diagram is self-explanatory and can be basically read like a family tree that is not rooted; branch lengths are proportional to linguistic distances. A long path therefore indicates many differences, while a short path indicates that the varieties are fairly similar. Sets of parallel lines and boxy shapes indicate splits in the data. Starting with the top section of Figure 3, we find the Philippines, Singapore, and Malaysia, together with the two metropolitan varieties, Great Britain and the United States. From this group, it is necessary to highlight the connection between the Philippines and the United States, since the Philippines is the only American colony included in this study. We should also point out Singapore within this group, since some signs of it becoming an L1 are visible (Buschfeld 2020a), which, together with the trend towards

the americanization of English (Buschfeld and Kautzsch 2017; Gilquin 2018; Gonçalves *et al.* 2018; Low and Pakir 2018), might explain its proximity to the United States. In terms of the substrate languages of this group, this figure confirms what was already discussed with Figure 2, that is, transfer of features from the substrate languages spoken in each region does not seem to be an explanatory factor. As can be seen in brackets next to each variety, the Philippines and Malaysia, both with finite complements in their substrates, are located near Singapore, which does not have clausal complementation. The varieties in this upper section of the figure are also in a mixture of phases of the Dynamic Model (Schneider 2007); Malaysia in phase 3, the Philippines and Singapore in phase 4, and the United States in phase 5. This also confirms that the evolutionary phase of development does not seem to explain the closeness of the varieties in the figure. On the other hand, if we look at the varieties in this group in terms of their geographical location, Singapore, Malaysia, and the Philippines are in what is commonly referred to as South-East Asia. Therefore, it seems that their geographical proximity may be behind their similarities.

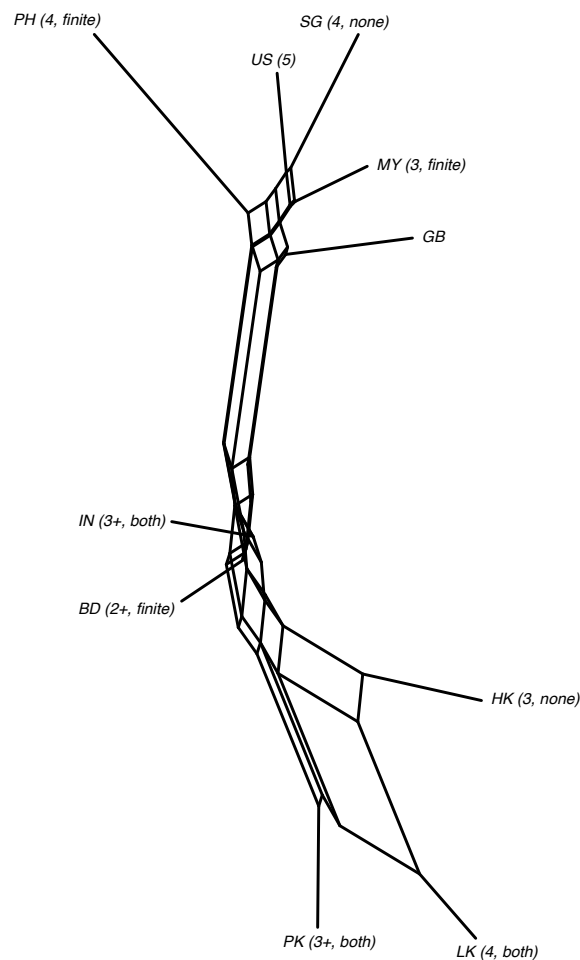


Figure 3: *NeighborNet* of similarity across Asian varieties of English



A look at the bottom section of the figure shows a similar picture. In this section, we find India, Bangladesh, Pakistan, Sri Lanka, and Hong Kong. First, it is important to highlight that there is an important historical connection between India, Pakistan, and Bangladesh that cannot be ignored; during the British Empire and, therefore, when English was introduced in the region, these three countries were one nation. However, regarding the Dynamic Model (Schneider 2007), these are in different phases. India and Pakistan are in an advanced stage of phase 3, while Bangladesh is still in phase 2. Additionally, in this group we also have Hong Kong in phase 3, and Sri Lanka in phase 4. Therefore, here the phase of development seems not to be sufficient to explain similarities and differences between varieties. As for the substrate languages, they do not seem to explain the proximity of the varieties since, in this group, there are English varieties with both finite and nonfinite complementation systems in their substrate languages (India, Pakistan, and Sri Lanka), one with only finite complements (Bangladesh), and one with no clausal complementation (Hong Kong). What does seem to explain the closeness between varieties, and therefore their similarities, is the geographical location. India, Bangladesh, Pakistan, and Sri Lanka are South Asian varieties, and Hong Kong, a little further away in the figure, is part of East Asia.

Therefore, from Figures 2 and 3, it can be concluded that the phase of development in Schneider's Dynamic Model (2007) and the transfer of features from the substrate languages—the major factors frequently studied in the literature as determinants of the variation in World Englishes—do not seem to account for the similarities and differences between the varieties of English studied here. If we look at the varieties individually, it may seem that these extra-linguistic factors could explain the preference for less complex structures within a non-categorical variation in ESL varieties. However, when studying a greater number of English varieties, it can be noticed that varieties in different evolutionary phases of development and with different complementation systems are similar in terms of their choice of less complex structures, which demonstrates that, at least in this case, these two factors are not as decisive as they may seem at first glance. On the contrary, a factor such as geographical location, which has not been studied very often and that cannot be taken into account with investigations of individual varieties, does seem to have a greater explanatory power of the similarities across English varieties.

## 5. CONCLUSION

This paper is a step forward in the study of CMC by analyzing the English used on the internet. The study analyzed more than 10,000 examples of the complementation of the verb REGRET on the GloWbE corpus in Asian varieties of English (India, Sri Lanka, Pakistan, Bangladesh, Singapore, Malaysia, the Philippines, and Hong Kong) and metropolitan varieties (Great Britain and the United States). There was a special focus on geographical proximity of the varieties as a potential extra-linguistic determining factor for the similarities and differences found, even though other factors frequently discussed in the literature—such as SLA and language contact processes, and the effect of the phase of evolution of the individual varieties in terms of Schneider’s Dynamic Model (2007)—were also considered.

The non-categorical variability with this verb is between finite *that*-clauses and nonfinite *-ing* clauses (*you will regret that you went to Lahore* vs. *you will regret going to Lahore*). Results showed a clear different distribution of these two patterns across World Englishes, with a general preference for *that*-clauses in ESL varieties, more specifically in India, Bangladesh, Hong Kong, Sri Lankan, and Pakistani Englishes. However, there are other three varieties in which a more frequent use of *-ing* clauses can be seen, in particular, Singapore, Malaysia, and the Philippines. The principle of maximization of transparency and the transfer of features from substrate languages, the extra-linguistic factors result of the SLA process, and the phase of development in Schneider’s Dynamic Model (2007), do not account for the similarities and differences between the varieties of English studied here. The non-hierarchical phylogenetic network (*NeighborNet*) has brought light to another extra-linguistic factor that has not often been studied in this area of linguistics, the geographical proximity of the varieties under research. The varieties displaying more similarities are those that are geographically close making the distinction between three geographical areas possible: South Asia (with India, Sri Lanka, Pakistan, and Bangladesh), South-East Asia (with Singapore, Malaysia, and the Philippines), and East Asia (Hong Kong).

The relevance of this study is that it has revealed the importance of the geographical location as a determining factor in the similarities and differences across World Englishes. The literature is not conclusive regarding this factor since there are studies that find it to be a weak predictor (Szmrecsanyi and Kortmann 2009b; Szmrecsanyi 2013) while in others it is the most important one (Kortmann and Schröter 2017; Fuchs *et al.* 2019). The

present investigation is another study that highlights the geographical proximity as the most important predictor. This study has also revealed the need to study larger sets of English varieties so that factors such as geographical proximity can be tested.

Future work should include a large variety of verbs so that the effect of the geographical proximity can be tested in the complementation system in general as well as focus on other linguistic features.

#### REFERENCES

- Andersen, Roger W. 1984. The one-to-one mapping principle of interlanguage construction. *Language Teaching* 34: 77–95.
- Ansari, Sahar and Sorin Draghici. 2019. *NeighborNet: Neighbor\_net Analysis. R Package* version 1.2.0.
- Bhatia, Tej K. 1993. *Punjabi: A Cognitive-Descriptive Grammar*. London: Routledge.
- Brezina, Vaclav and Miriam Meyerhoff. 2014. Significant or random? A critical review of sociolinguistic generalizations based on large corpora. *International Journal of Corpus Linguistics* 19: 1–28.
- Brugmann, Karl. 1909. Das Wesen der lautlichen Dissimilationen. *Abhandlungen der philologisch-historischen Klasse der königlich-sächsischen Gessellschaft der Wissenschaften* 27: 141–178.
- Brunner, Thomas. 2014. Structural nativization, typology and complexity: Noun phrase structures in British, Kenyan and Singaporean English. *English Language and Linguistics* 18/1: 23–48.
- Brunner, Thomas. 2017. *Simplicity and Typological Effects in the Emergence of New Englishes: The Noun Phrase in Singaporean and Kenyan English*. Berlin: Mouton de Gruyter.
- Bryant, Davis and Vincent Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution* 21/2: 255–265.
- Buschfeld, Sarah. 2020a. *Children's English in Singapore: Acquisition, Properties, and Use*. London: Routledge.
- Buschfeld, Sarah. 2020b. Language acquisition and World Englishes. In Daniel Schreier, Marianne Hundt and Edgar W. Schneider eds., 559–584.
- Buschfeld, Sarah and Alexander Kautzsch. 2017. Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes* 36/1: 104–126.
- CIA. 2024. CIA: *The World Factbook*. <https://www.cia.gov/the-world-factbook/countries/> (22 March, 2024.)
- Cuyckens, Hubert, Frauke D'hoedt and Benedikt Szmrecsanyi. 2014. Variability in verb complementation in Late Modern English: Finite vs. non-finite patterns. In Marianne Hund ed. *Late Modern English Syntax*. Cambridge: Cambridge University Press, 182–204.
- Cysouw, Michael. 2007. New approaches to cluster analysis of typological indices. In Peter Grzybek and Reinhard Köhler eds. *Exact Methods in the Study of Language and Text*. Berlin: Mouton de Gruyter, 61–76.
- Davies, Mark. 2012. Some methodological issues related to corpus-based investigations of recent syntactic changes in English. In Terttu Nevalainen and Elizabeth C.

- Traugott eds. *The Oxford Handbook of the History of English*. Oxford: Oxford University Press, 157–174.
- Davies, Mark and Robert Fuchs. 2015a. 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.
- Davies, Mark and Robert Fuchs. 2015b. A reply. *English World-Wide* 36/1: 45–47.
- Deshors, Sandra C. and Stefan T. Gries. 2016. Profiling verb complementation constructions across New Englishes. *International Journal of Corpus Linguistics* 21/2: 192–218.
- Deshors, Sandra C. and Tobias Bernaisch. 2019. Corpus approaches to World Englishes: A bird's-eye view. In Peter I. De Costa, Dustin Crowther and Jeffrey Maloney eds. *Investigating World Englishes: Research Methodology and Practical Implications*. London: Routledge, 21–43.
- Fang, Meili. 2010. *Spoken Hokkien*. London: University of London.
- Fuchs, Robert, Bertus van Rooy and Ulrike Gut. 2019. Corpus-based research on English in Africa: A practical introduction. In Alexandra U. Esimaje, Ulrike Gut and Bassey E. Antia eds. *Corpus Linguistics and African Englishes*. Amsterdam: John Benjamins, 37–69.
- García-Castro, Laura. 2018. *The Complementation Profile of REMEMBER in Post-colonial Englishes*. Vigo: University of Vigo dissertation.
- Gilquin, Gaëtanelle. 2018. American and/or British influence on L2 Englishes – Does context tip the scale(s)? In Sandra C. Deshors ed. *Modeling World Englishes: Assessing the Interplay of Emancipation and Globalization of ESL Varieties*. Amsterdam: John Benjamins, 187–216.
- Givón, Thomas. 1985. Iconicity, isomorphism and non-arbitrary coding in syntax. In John Haiman ed. *Iconicity in Syntax*. Amsterdam: John Benjamins, 187–219.
- Gonçalves, Bruno, Lucía Loureiro-Porto, José J. Ramasco and David Sánchez. 2018. Mapping the americanization of English in space and time. *PLoS ONE* 13/5: 1–15.
- Gries, Stefan T. and Tobias Bernaisch. 2016. Exploring epicentres empirically: Focus on South Asian Englishes. *English World-Wide* 37/1: 1–25.
- Gupta, Anthea F. 1994. *The Step-Tongue. Children's English in Singapore*. Clevedon: Multilingual Matters.
- Gut, Ulrike. 2011. Studying structural innovations in New English varieties. In Joybrato Mukherjee and Marianne Hundt eds. *Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam: John Benjamins, 166–205.
- Haiman, John. 1985. *Natural Syntax*. Cambridge: Cambridge University Press.
- Haspelmath, Martin, König Ekkehard, Wulf Oesterreicher and Wolfgang Raible. 2001. *Language Typology and Language Universals: An International Handbook*. Berlin: Walter de Gruyter.
- Heller, Benedikt and Melanie Röthlisberger. 2015. Big data on trial: Researching syntactic alternations in GloWbE and ICE. Paper presented at the *Data to Evidence (d2e) Conference*. Helsinki: University of Helsinki, 21 October 2015.
- Herring, Susan C. 1996. *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*. Amsterdam: John Benjamins.
- Hundt, Marianne. 2013. The diversification of English: Old, new and emerging epicenters. In Daniel Schreier and Marianne Hundt eds. *English as a Contact Language*. Cambridge: Cambridge University Press, 182–203.
- Hundt, Marianne. 2020. Corpus-based approaches to World Englishes. In Daniel Schreier, Marianne Hundt and Edgar W. Schneider eds., 506–533.

- Kachru, Brag B. ed. 1982. *The Other Tongue: English across Cultures*. Urbana: University of Illinois Press.
- Karmiloff-Smith, Annette. 1979. *A Functional Approach to Child Language*. Cambridge: Cambridge University Press.
- Kortmann, Bernd and Verena Schröter. 2017. Varieties of English. In Raymond Hickey ed. *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press, 304–330.
- Koul, Omkar N. 2008. *Modern Hindi Grammar*. Springfield: Dunwoody Press.
- Krishnamurti, Bhadriraju and John Peter Lucius Gwynn. 1985. *A Grammar of Modern Telugu*. New Delhi: Oxford University Press.
- Lehmann, Thomas. 1993. *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Levshina, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins.
- Lim, Lisa. 2017. Southeast Asia. In Markku Filppula, Juhani Klemola and Sharma Devyani eds. *The Oxford Handbook of World Englishes*. Oxford: Oxford University Press, 448–471.
- Lim, Lisa and Joseph A. Foley. 2004. English in Singapore and Singapore English: Background and methodology. In Lisa Lim ed. *Singapore English: A Grammatical Description*. Amsterdam: John Benjamins, 1–18.
- Low, Ee-Ling and Anne Pakir. 2018. *World Englishes: Rethinking Paradigms*. London: Routledge.
- 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.
- Matthews, Stephen and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- McMahon, April M. S. and Robert McMahon. 2005. *Language Classification by Numbers*. Oxford: Oxford University Press.
- McMahon, April M. S., Paul Heggarty, Robert McMahon and Warren Maguire. 2007. The sound patterns of Englishes: Representing phonetic similarity. *English Language and Linguistics* 11: 113–142.
- Mufwene, Salikoko S. and Stefan T. 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. 2015. Responses to Davies and Fuchs. *English World-Wide* 36/1: 34–37.
- Mukherjee, Joybrato and Sebastian Hoffmann. 2006. Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27/2: 147–173.
- Mukherjee, Joybrato and Stephan T. Gries. 2009. Collostructional nativisation in New Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide* 30/1: 27–51.
- Nelson, Gerald. 2015. Responses to Davies and Fuchs. *English World-Wide* 36/1: 38–40.
- Nordhoff, Sebastian. 2009. *A Grammar of Upcountry Sri Lanka Malay*. Utrecht: LOT.
- Omar, Asmah H. and Rama Subbiah. 1989. *An Introduction to Malay Grammar*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Pandharipande, Rajeshwari V. 1997. *Marathi*. London: Routledge.
- Peters, Pam and Tobias Bernaisch. 2022. The current state of research into linguistic epicentres. *World Englishes* 41: 320–332.

- Platt, John, Heidi Weber and Ho Mian Lian. 1984. *The New Englishes*. London: Routledge and Kegan Paul.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7/2: 149–182.
- Rohdenburg, Günter. 2003. Cognitive complexity and horror aequi as factors determining the use of interrogative clause linkers in English. In Günter Rohdenburg and Britta Mondorf eds. *Determinants of Grammatical Variation in English*. Berlin: Mouton de Gruyter, 205–250.
- 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.
- Romasanta, Raquel P. 2019. Variability in verb complementation: Determinants of grammatical variation in indigenized L2 varieties of English. In Hanna Parviainen, Mark Kaunisto and Päivi Pahta eds. *Corpus Approaches into World Englishes and Language Contrasts*. Helsinki: VARIENG. <https://varieng.helsinki.fi/series/volumes/20/romasanta/>
- Romasanta, Raquel P. 2021. Substrate language influence in Postcolonial Asian Englishes and the role of transfer in the complementation system. *English Studies* 102/8: 1151–1170.
- Santoro, Gerald M. 1995. What is computer-mediated communication? In Mauri P. Collins and Zane L. Berge eds. *Computer Mediated Communication and the Online Classroom*. Cresskill: Hampton, 11–27.
- Schachter, Paul and Fe T. Otnes. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Schiffman, Harold F. 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 2012a. 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. 2012b. Exploring the interface between World Englishes and second language acquisition – and implications for English as lingua franca. *Journal of English as a Lingua Franca* 1/1: 57 – 91.
- Schneider, Edgar W. 2013. English as a contact language: The ‘New Englishes’. In Daniel Schreier and Marianne Hundt eds. *English as a Contact Language*. Cambridge: Cambridge University Press, 131–148.
- Schreier, Daniel, Marianne Hundt and Edgar W. Schneider eds. 2019. *The Cambridge Handbook of World Englishes*. Cambridge: Cambridge University.
- Setter, Jane, Cathy S. P. Wong and Brian H. S. Chan. 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.
- Slobin, Dan. 1973. Cognitive prerequisites for the development of grammar. In Charles Ferguson and Dan Slobin eds. *Studies in Child Language Development*. New York: Holt, Rinehard and Winston, 175–208.
- Slobin, Dan. 1977. Language change in childhood and history. In J. Macnamara ed. *Language Thought and Language Learning*. New York: Academic Press, 185–214.

- 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*. Winheim: Verlag Chemie, 229–243.
- Steger, Maria. 2012. *New Englishes are Simpler, isn't it? Morphosyntactic Iconicity in Institutionalized Second-Language Varieties of English*. Regensburg: University of Regensburg dissertation.
- Steger, Maria and Edgar W. 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, Indigenization, Contact*. Berlin: Mouton de Gruyter, 156–191.
- Stevens, Peter. 1980. *Teaching English as an International Language: From Practice to Principle*. Oxford: Pergamon Press.
- Szmrecsanyi, Benedikt. 2013. Typological profile: L1 varieties. In Bernd Kortmann and Kerstin Lunkenheimer eds. *The Mouton World Atlas of Variation in English*. Berlin: Mouton de Gruyter, 826–843.
- Szmrecsanyi, Benedikt and Bernd Kortmann. 2009a. Between simplification and complexification: Non-standard varieties of English around the world. In Geoffrey Sampson, David Gil and Peter Trudgill eds. *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press, 64–79.
- Szmrecsanyi, Benedikt and Bernd Kortmann. 2009b. The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119/11: 1643–1663.
- Szmrecsanyi, Benedikt and Christoph Wolk. 2011. Holistic corpus-based dialectology. *Brazilian Journal of Applied Linguistics* 11/2: 561–592.
- Szmrecsanyi, Benedikt and Jason Grafmiller. 2023. *Comparative Variation Analysis: Grammatical Alternations in World Englishes*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt and Melanie Röthlisberger. 2019. World Englishes from the perspective of dialect typology. In Daniel Schreier, Marianne Hundt and Edgar W. Schneider eds., 534–558.
- Tan, Ying-Ying. 2014. English as a ‘mother tongue’ in Singapore. *World Englishes* 33/3: 319–339.
- Tegey, Habibullah and Barbara Robson. 1996. *A Reference Grammar of Pashto*. Washington: Department of Education.
- Thomason, Sarah G. 2001. *Language Contact: An Introduction*. United States: Mouton de Gruyter.
- Werner, Valentin. 2014. *The Present Perfect in World Englishes: Charting Unity and Diversity*. Bamberg: University of Bamberg Press.
- Wheeler, Benjamin, Robert Englebretson and Carol Genetti eds. 2005. *Complementation in Colloquial Sinhala: Observations on the Binding Hierarchy*. Santa Barbara: University of California, Santa Barbara.
- Williams, Jessica. 1987. Non-native varieties of English: A special case of language acquisition. *English World-Wide* 8/2: 161–199.
- Yegorova, Raisa Petrovna. 1971. *The Sindhi Language*. Moscow: Nauka Publishing House.

*Corresponding author*

Raquel P. Romasanta  
University of Santiago de Compostela  
Facultade de Filoloxía  
Avda. de Castelao, s/n  
15782 Santiago de Compostela  
Spain  
E-mail: [raquel.romasanta@usc.es](mailto:raquel.romasanta@usc.es)

received: November 2023

accepted: February 2024