

# Clausal Verb Complementation in Varieties of English

Laetitia Van Driessche (University of Zurich) and Hubert Cuyckens (University of Leuven)<sup>1</sup>

laetitia.vandriessche@es.uzh.ch and hubert.cuyckens@kuleuven.be

---

## Abstract

*This article looks at the finite/non-finite complementation alternation with expect and suggest (in its suasive meaning) from a diachronic and a synchronic perspective. It investigates whether the diachronic shift from finite to non-finite complementation in British English is reflected by distributional differences in finite vs. non-finite complementation patterns in World Englishes positioned at different stages in Schneider's Dynamic Model (2007). It also examines the factors that determine the complementation alternation in these varieties. Data have been extracted from the CLMET, BNC and GloWbE corpora. Methodologically, frequency analyses, random forest analyses, logistic regression analyses and conditional inference trees are employed. The analyses show that expect largely corroborates the hypotheses, whereas suggest shows unexpected tendencies.*

---

## 1 Introduction

This article deals with the variation of finite vs. non-finite complement clauses (CCs) with the complement-taking predicates (CTPs) *expect* and *suggest*. In the present study, finite complementation involves *that*-clauses (1a, 2a), whereas non-finite complementation comprises gerundial (2b) and infinitival clauses (1b, 2c).

- (1) a. *She had half expected that he would deflect the question.* (BNC)  
b. *I [...] cannot expect the bookseller to take them again.* (CLMET)
- (2) a. *Security experts suggest that you steer clear of these bogus pages.* (GloWbE BrE)  
b. *[H]e suggested bringing his mum.* (GloWbE BrE)  
c. *I would suggest to bring Android apps out first.* (GloWbE BrE)

The study will be restricted to CCs occurring *after* the CTP. Fronted CCs will not be considered here as it has been observed that processing works differently in those contexts (see Hawkins 1990).

Two research questions are addressed in this article. The first is diachronic as well as synchronic in nature. Diachronically, it has been observed that an increasing share of finite clauses has been replaced by non-finite clauses in British English (see Rohdenburg 1995; Denison 1998; Los 2005; Cuyckens/D'Hoedt/Szmrecsanyi 2014). Considering that finite CCs can be viewed as more isomorphic-iconic than non-finite CCs because they tend to display a one-to-one correspondence between meaning and form (they overtly express subjects, tense, aspect and modality (cf. Noonan 2007: 59)), clausal complementation can be said to have become less iconic, or more economic, over time. Synchronically, it has been argued that iconicity facilitates second-language acquisition, and by extension the development of second language (L2)

---

<sup>1</sup> We would like to thank Benedikt Szmrecsanyi and the two anonymous reviewers for their valuable feedback on this paper.

varieties in Schneider's (2007) Dynamic Model (Steger/Schneider 2012: 158–159, 164). Accordingly, it can be hypothesised that varieties of English at a less advanced stage in the Dynamic Model are likely to contain more iconic structures, i.e. finite CCs, than varieties at a more advanced stage (Steger/Schneider 2012: 164). Combining the diachronic and synchronic perspectives, this paper will investigate whether the diachronic shift from finite to non-finite complementation is reflected by varieties positioned at increasingly more advanced stages in the Dynamic Model. For this purpose, a frequency study will be carried out, comparing distributions of finite vs. non-finite complements across diachronic stages of British English with those of the varieties under investigation.

The second goal of this paper is to identify the factors that significantly determine complementation alternation in World Englishes. Special attention will be given to the strength of the factor 'second language variety' in predicting finite vs. non-finite clausal complementation in that we will try to ascertain whether there is indeed an L2 learning effect across the World Englishes and whether this learning effect is related to the level of advancedness in the Dynamic Model.<sup>2</sup> This will be examined with random forest analyses, logistic regression analyses and conditional inference trees. It will become clear that there are important differences between *expect* and *suggest*: whereas *expect* seems to conform to the hypotheses, *suggest* tells quite a different story.

## 2 Data and Methodology

The article will focus on two CTPs: *expect* and *suggest* in its suasive meaning.<sup>3</sup> The patterns for *expect* consist of infinitival CCs (non-finite) and *that*-clauses (finite) in the indicative (often with *will/would*) or the subjunctive. For *suggest*, these consist of infinitival and gerundial CCs (non-finite) and *that*-clauses (finite) in the indicative, the subjunctive or with mandative *should*. Modals different from suasive (or mandative) *should* were not included as it could be argued that these add a nuance that cannot be expressed by the non-finite complements.<sup>4</sup> Infinitival clauses following *suggest* are the proscribed variant (cf. Carter et al. 2016, "Suggest"). However, looking into possible influence from prescriptivism in the different varieties under study lies beyond the scope of this paper.

Zero-*that*-clauses were not included for either verb as they pose problems for recall: unlike for the other clause types, they do not contain an easily retrievable element (e.g. the *that*-complementiser).<sup>5</sup>

The patterns were extracted from the CLMET3.0 (Late Modern English (LModE)) and the BNC (Present-day British English (PDE)) to trace their relative distribution over time, and from GloWbE (World Englishes) to examine their frequency distribution across L2 varieties. The BNC data were restricted to a sample from the fiction subsection to maintain genre conformity with the CLMET.

---

<sup>2</sup> Investing the link between World and Learner Englishes with regard to this second language acquisition effect falls outside the scope of this paper. A useful study on clausal complementation in Learner Englishes is Tizón-Couto (2014).

<sup>3</sup> In its non-suasive meaning, there is no alternation with the gerund.

<sup>4</sup> For instance, in (i) the speaker moderates the suasive nature of the suggestion by using the modal *could* instead of *should*.

(i) *I have suggested over the years that more co-operatives and friendly societies could be set up for employment.* (GloWbE JamE)

The non-finite correlate of (i) would lose the moderating nuance.

<sup>5</sup> For a description of the regular expressions used to extract the clauses, see Van Driessche (2018).

The five varieties selected to represent the different phases in Schneider's Dynamic Model are British (reference variety), Canadian (phase 5), Jamaican (phase 4), Malaysian (phase 3) and Bangladeshi English (phase 2).<sup>6</sup> There is no phase 1 variety as this consists of the settlement of a new colony (Schneider 2007: 33-36), which means that it cannot be investigated for PDE.

To examine which factors determine the variation of finite vs. non-finite CCs in World Englishes, data extracted from GloWbE were coded for a number of factors. Coded data were then entered into random forest and logistic regression analyses and into conditional inference trees. Random forest analyses look at *factor importance*: they measure the factors' "relative importance" in the choice between two alternatives (Tagliamonte/Baayen 2012: 159–160, 172). Logistic regression analyses look at *factor strength*: the effect that the presence of a factor has on the response variable (in this case finite vs. non-finite complementation) (Tagliamonte/Baayen 2012: 150; Speelman 2014: 501–502). Conditional inference trees show how factors interact with each other and can therefore be used to find probabilistic grammar differences between varieties (see Szmrecsanyi et al. 2016).

In the following paragraphs (2.1-2.4), we discuss the factors used in the statistical analyses. The labels used for the factors are provided between brackets.

## 2.1 Factors Relating to the Cognitive Complexity Principle

As we already mentioned, the choice between finite and non-finite complementation also entails a choice between more or less explicit grammatical options. In this regard, it has been claimed by Rohdenburg (1996: 151) that "the more explicit [option] will tend to be favored in cognitively more complex environments"; this is Rohdenburg's Complexity Principle. His principle is here operationalised by the following six factors.

A first type of a cognitively complex environment (Rohdenburg 1996: 160–161) involves the **Intervening material** between the CTP and the complementiser (**Intervening material**), measured in terms of the number of words between the CTP and the *that*-complementiser (one word in (3a)), the CTP and the gerund (two words in (3b)), the CTP and the infinitive marker *to* (three words in (3c)), or between the infinitive marker *to* and the infinitive itself (one word in (3d)). Not considered intervening material are raised objects preceding *to*-infinitives with the CTP *expect* (such as *it* in (4a)), because for their finite counterparts the object would be part of the *that*-clause (i.e. would not precede the *that*-complementiser) (4b).

(3) a. [...] *I suggested then that the Government should approach a programme of publicizing [...]* (GloWbE JamE)

b. *Bryant suggested to Watson writing a private letter [...]* (GloWbE BrE)

c. *I have already suggested to my son to go there [...]* (GloWbE BrE)

d. *We've suggested to them to just make the proposal.* (GloWbE CanE)

(4) a. *Don't expect it to be easy, but give it a try [...]* (GloWbE BrE)

b. *Don't expect that it will be easy, but give it a try [...]*

---

<sup>6</sup> The only phase 2 variety that Schneider (2007) mentions for his Dynamic Model (i.e. Fiji English) is not represented in GloWbE. For this reason, BanE was chosen to represent this phase after a perusal of literature on the status of this variety.

For the same reason, a negator preceding the gerund or the *to*-complementiser (5a), or positioned between the *to*-complementiser and the infinitive (5b), is not considered intervening material either.

- (5) a. *So, do you honestly suggest me not to refer to the Daily Star to improve my English?* (GloWbE BanE)  
b. *I suggest to not join [...]* (GloWbE BanE)

As this is a continuous variable (as are Clause length and Formality, see below), the regression analysis calculates the likelihood that one of the complementation patterns will be chosen when the number of words between the CTP and the complementiser increases (cf. Levshina 2015: 261). Counting for the continuous variables was done automatically in R.

A second complexity environment is **Clause length (Clause\_length)** (Rohdenburg 1996: 164), which is measured in terms of the number of words from the CTP until the end of the sentence (seven in (6)).

- (6) *[A]sk the social worker's supervisor to review the situation or suggest that a mediator be brought in.* (GloWbE CanE)

A third and fourth type of complex environments are **Negation of the CTP (Negation\_CTP)** (Rohdenburg 1996: 164) and **Negation in the subordinate clause (Negation\_sub)** (Rohdenburg 2015). Markers of negation comprise *not* (5a-b, 7a), *n't* (7b), *nor* (7c) *neither* (7d) and *not that* (7e).

- (7) a. *I am not suggesting that we stop having political debates.* (GloWbE MalE)  
b. *I don't suggest doing any serious physical activity [...]* (GloWbE JamE)  
c. *[...] I wasn't disappointed, nor was I expecting to be.* (GloWbE JamE)  
d. *It neither suggests that school gates are locked to prevent students leaving at lunchtime or that pupils are forced to eat specific foods.* (GloWbE BrE)  
e. *Not that I really expected to escape so easily.* (GloWbE MalE)

Fifth, **Passive CTPs (Passive)** are considered to be cognitively complex relative to their active counterparts (Rohdenburg 1996: 169, 173). Only attestations including *expect* (8a-b) are coded for this factor, since the gerund (a complementation pattern for *suggest*) does not usually combine with a CTP in the passive voice.

- (8) a. [...] *it* is expected that *hats be worn during prayer*. (GloWbE CanE)  
 b. *Buddhists* are expected to *know the difference between right and wrong* [...] (GloWbE CanE)

A final factor is **Coreferentiality (Coreferentiality)**, which, like Passive, will only be coded for *expect*. The subjects of the main and the subordinate clause can be coreferential, as in (9a), or refer to two different entities, as in (9b). Non-coreferentiality is cognitively more complex (Steger/Schneider 2012: 169-171).

- (9) a. *How can you expect to have an honest relationship if you do such stunts* [...] (GloWbE BrE)  
 b. Don't expect it to be *easy, but give it a try* [...] (GloWbE BrE)

## 2.2 Formality: number of contractions (Contractions)

The degree of **Formality** may also have an effect on the variation between the complementation types. Since it has been claimed that more formal registers prefer more explicit coding (cf. Rohdenburg 1996: 160), this would imply more finite complementation in formal contexts.

As a measure of the text's formality, AntConc's window size was set at 1000 characters<sup>7</sup> starting with *expect/suggest* (i.e. the right context), after which the number of contractions (following Quirk et al. 1985: 123) was counted in R.

## 2.3 Person and Tense of the CTP (Person\_CTP; Tense\_CTP)

Because we wanted to test the influence of deictic distance<sup>8</sup> on the complementation type, we also coded for **Person (Person\_CTP) and Tense of the CTP (Tense\_CTP)**. The **Person of the CTP** is coded for the first, second and third person. This factor was added because it proved to be a significant factor in Cuyckens/D'Hoedt's (2015: 89-94) analysis of complementation patterns of the verb *admit*, although with conflicting results. When *that*-clauses and zero-*that*-clauses were contrasted with gerundial, *to*-gerundial and *to*-infinitival clauses, their conditional inference tree showed that the third person predicted non-finite complementation in PDE when *admit* had the sense of *wrongdoing*. However, when *that*-clauses were contrasted with *to*-infinitival clauses and zero-*that*-clauses, the third person favoured *that*-clauses.<sup>9</sup>

As the future tense is not frequent, coding for **Tense of the CTP** is restricted to a 'past' and a 'non-past' level. While 'past' obviously comprises situations occurring in a past time period (including situations encoded by modals with past time reference, as in (10a)), 'non-past' comprises CTPs in the present and future tense as well as modals expressing a hypothetical situation such as *can* in (10b) or *could* in (10c).

<sup>7</sup> There were minimal variations in the window size that AntConc provided.

<sup>8</sup> The third person and the past tense are more deictically removed from the speaker's "zero-point [...] of the deictic context" (Lyons 1977: 683).

<sup>9</sup> There are two different combinations of clausal complementation in this article since non-coreferentiality between the subject of the main and the subordinate clause does not allow for all complementation patterns of *admit* (Cuyckens/D'Hoedt 2015: 78, 87).

- (10) a. *That was very common in England just 100 years ago. Women could not necessarily expect to survive childbirth.* (GloWbE MaleE)
- b. *By the end of 2011, we can expect the figure to reach RM450bn.* (GloWbE MaleE)
- c. *With its mass release in FTV sets, you could expect to see its price fall by quite a bit once it is released [...]* (GloWbE MaleE)

## 2.4 Variety (Variety)

As the present study probes the effect of the factor ‘second language variety’ on finite vs. non-finite clausal complementation, each of the attestations will be coded for **Variety** (as mentioned above, these are British (BrE), Canadian (CanE), Jamaican (JamE), Malaysian (MalE) and Bangladeshi English (BanE)).

## 3 Results

In what follows, the results for *expect* (3.1) and *suggest* (3.2) will be described. These results will then be discussed in section 4.

### 3.1 *Expect*

The first section (3.1.1) provides the frequency distributions for *expect*; the second section (3.1.2) presents the results of the statistical analyses.

#### 3.1.1 Frequency distributions

Table 1 presents the frequency distributions for finite vs. non-finite complementation in the CLMET (1780-1850) and the BNC (1980-1993). The results from the CLMET are confined to a sample from the second subperiod of the corpus (i.e. the middle subperiod) as there were sufficient attestations for analysis in this subperiod (as opposed to *suggest*, which was less frequent).

Table 1: Frequencies and percentages for *expect* in the CLMET (1780-1850) and BNC (1980-1993); normalised per 100,000 words.

Corpus	Complement type	Raw frequencies	Percentages	Normalised frequencies
CLMET	Finite	54	16.5%	1.8
	Non-finite	274	83.5%	9.4
BNC	Finite	11	2.8%	0.4
	Non-finite	385	97.2%	12.4

Table 1 indicates that there is a significant rise in non-finite complementation between the periods 1780-1850 and 1980-1993 (a chi-square test indicates that  $p = 1.43263e-10$ ). In fact, the BNC consists almost exclusively of non-finite clauses (97.2%).

Table 2 presents the frequencies for *expect* in a sample of the GloWbE corpus (General section).

Table 2: Frequencies and percentages for *expect* in the GloWbE corpus; normalised per 100,000 words.

Variety	Complement type	Raw frequencies	Percentages	Normalised frequencies
British English	Finite	50	5.7%	0.8
	Non-finite	831	94.3%	13.6
Canadian English	Finite	56	7.2%	1.0
	Non-finite	719	92.8%	12.8
Jamaican English	Finite	53	9.7%	1.3
	Non-finite	495	90.3%	12.5
Malaysian English	Finite	52	5.1%	0.7
	Non-finite	960	94.9%	13.0
Bangladeshi English	Finite	50	11.1%	1.2
	Non-finite	400	88.9%	9.4

Table 2 shows that non-finite complementation is highly frequent in all varieties. The highest percentage occurs in MalE (94.9%), followed by BrE (94.3%), CanE (92.8%), JamE (90.3%) and BanE (88.9%).

Chi-square analyses show that apart from JamE ( $p = 0.456986305$ ), all varieties differ significantly from BanE. The difference between JamE and BrE (0.004508678) and JamE and MalE ( $p = 0.000646521$ ) is also significant. The differences between the other varieties are not significant, however.

### 3.1.2 Statistical analyses

The random forest analysis<sup>10</sup> (Figure 1) shows the relative importance of each factor in explaining the variation between finite and non-finite complementation. Coreferentiality explains the alternation best, followed by the factors Variety, Passive CTP, Person of the CTP and Clause length. Intervening material and Tense of the CTP are not important factors. The index of concordance for this analysis is  $C = 0.8499125$ , which means that it has a good fit (Tagliamonte/Baayen 2012: 156).

<sup>10</sup> The default settings were used for the random forest analysis.

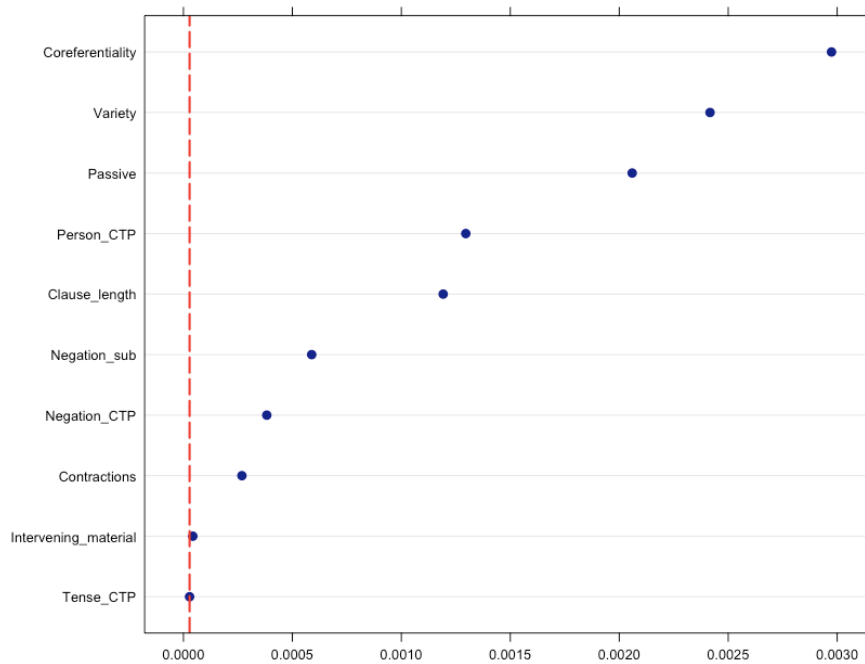


Figure 1: Random forest analysis for *expect*.  $C = 0.8499125$ .

The logistic regression analysis in Figure 2 provides information about the strength of each factor in predicting the alternation. The predicted output is for non-finite complementation. A negative value for ‘Estimate’ means that the factor level (e.g. BanE for the factor Variety) disfavors non-finite complementation; a positive value indicates a favouring effect. The effect is always measured against a reference level, e.g. BrE for Variety. The reference level is not visible in the model (Tagliamonte/Baayen 2012: 148; Speelman 2014: 521-522).

The final two columns show the significance of the (dis)favouring effect. The boundary for significance is  $p = 0.05$  (Speelman 2014: 507-508). The degree of significance is represented more graphically in the final column: if there are no stars in the final column, the effect is not significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	2.883351	0.233417	12.353	< 2e-16	***
VarietyBanE	-0.823640	0.219640	-3.750	0.000177	***
VarietyCanE	-0.349547	0.208419	-1.677	0.093516	.
VarietyJamE	-0.538444	0.213778	-2.519	0.011779	*
VarietyMale	0.134871	0.210490	0.641	0.521684	
PassiveYes	0.540264	0.163897	3.296	0.000979	***
Person_CTP2nd	0.484121	0.261260	1.853	0.063879	.
Person_CTP3rd	0.224733	0.177212	1.268	0.204740	
Negation_CTPYes	0.090566	0.221266	0.409	0.682313	
Negation_subYes	-2.269328	0.422248	-5.374	7.68e-08	***
Tense_CTPPast	-0.211283	0.169810	-1.244	0.213414	
CoreferentialitySame	2.431745	0.369332	6.584	4.57e-11	***
Clause_length	-0.037779	0.005667	-6.666	2.62e-11	***
Intervening_material	-0.074183	0.063514	-1.168	0.242817	
Contractions	0.145561	0.078453	1.855	0.063541	.

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Figure 2: Logistic regression analysis for *expect*. The predicted output is for non-finite complementation.



The logistic regression analysis confirms the frequency count: BrE has a slightly stronger tendency to favour non-finite complementation, except in the case of MalE (although this favouring effect is not significant). The analysis also shows that the passive voice, the second person, coreferentiality and informality have a favouring effect on non-finite complementation, whereas negation in the subordinate clause and longer CCs have a disfavouring effect.

The conditional inference tree in Figure 3 chooses the factor that “has the strongest association with the response” (i.e. response variable), after which it “makes a binary split in this variable, dividing the dataset into two subsets” (Levshina 2015: 291). These steps are then repeated “until there are no variables that are associated with the outcome at the pre-defined level of statistical significance” (Levshina 2015: 291). The tree indicates that Clause Length is the most predictive factor (Node 1): CCs of less than or equal to 16 words strongly predict non-finite complementation (Nodes 4-6). In the absence of coreferentiality, longer clauses predict non-finite complementation slightly less than shorter clauses (Nodes 4 and 5). When Clause Length is higher than 16 words, there is a high predictive effect for non-finite complementation when the two subjects are coreferential (Node 19). When there is no coreferentiality, Negation in the Subordinate Clause is a predictive factor (Node 8), with negation predicting finite complementation (Node 18). Variety appears at a low branching level (Node 9), which indicates that it is only a reliable factor when there is no negation in the subordinate clause, when there is no coreferentiality, and when the clause is longer than 16 words. The varieties pattern differently: for BanE, CanE and JamE, Passive CTPs (Node 10), and, in the case of an active CTP, Negation of the CTP (Node 11) are predictive factors. For BrE and MalE, Intervening Material is a predictive factor (Node 15): there is a higher predictive effect for non-finite complementation in contexts with no intervening material (Nodes 16 and 17).

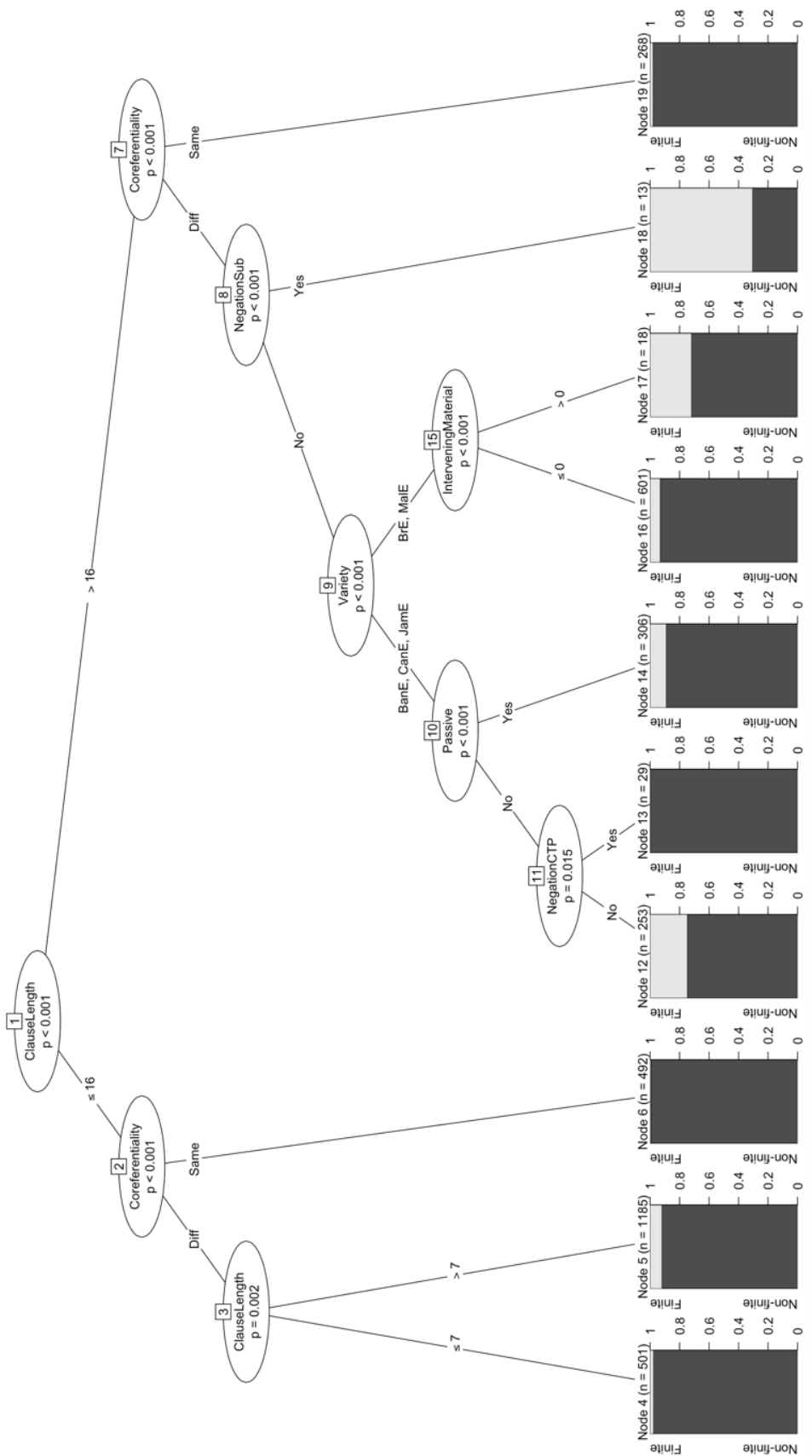


Figure 3. Conditional inference tree for *expect*.

### 3.2 Suggest

The following sections provide the results of the frequency distributions (3.2.1) and the statistical analyses (3.2.2) for *suggest* in its suasive meaning.

#### 3.2.1 Frequency distributions

Table 3 shows the frequency distributions for suasive *suggest* in the CLMET and the BNC. Due to the low frequencies for this verb, all subperiods of the CLMET are included (covering the period 1710-1920).

Table 3. Frequencies and percentages for *suggest* in the CLMET and BNC; Normalised per 1,000,000 words.

Corpus	Complement type	Raw frequencies	Percentages	Normalised frequencies
1710-1780 (CLMET)	Finite	0	0%	0
	Non-finite	0	0%	0
1780-1850 (CLMET)	Finite	16	76.2%	1.4
	Non-finite	5	23.8%	0.4
1850-1920 (CLMET)	Finite	101	78.3%	7.9
	Non-finite	28	21.7%	2.2
1980-1993 (BNC)	Finite	90	75%	10.5
	Non-finite	30	25%	3.5

Table 3 shows that there is a slight increase in non-finite complementation when the periods 1780-1850 and 1980-1993 and the periods 1850-1920 and 1980-1993 are compared. However, the difference between these periods is not significant ( $p = 0.907252532$  and  $p = 0.538855245$  respectively). In fact, none of the periods differ significantly from each other.<sup>11</sup>

The frequencies for *suggest* in a sample of the GloWbE corpus (General section) are displayed in Table 4.

---

<sup>11</sup> The p-value for the periods 1780-1850 and 1850-1920 is  $p = 0.829099237$ .

Table 4. Frequencies and percentages for *suggest* in the GloWbE corpus; Normalised per 1,000,000 words.

Variety	Complement type	Raw frequencies	Percentages	Normalised frequencies
British English	Finite	203	62.7%	13.6
	Non-finite	121	37.3%	8.1
Canadian English	Finite	244	54.3%	14.5
	Non-finite	205	45.7%	12.2
Jamaican English	Finite	233	80.1%	18.0
	Non-finite	58	19.9%	4.5
Malaysian English	Finite	170	50%	12.8
	Non-finite	170	50%	12.8
Bangladeshi English	Finite	163	48.4%	12.1
	Non-finite	174	51.6%	12.9

The results in Table 4 show that MalE and BanE have the highest share of non-finite clauses, with an even distribution for MalE and more non-finite than finite complementation for BanE.<sup>12</sup> The other three varieties have a higher proportion of finite complementation. Chi-square tests indicate that the differences between CanE and MalE ( $p = 0.226398397$ ), between CanE and BanE ( $p = 0.097095452$ ) and between BanE and MalE are not significant ( $p = 0.671056858$ ). In all other cases, the varieties differ significantly from each other.

### 3.2.2 Statistical analyses

The random forest analysis for *suggest* (Figure 4) shows that Intervening Material is the most important factor in the choice between finite and non-finite complementation, followed by Variety, Clause length and Formality. Tense of the CTP, Person of the CTP and Negation are less important factors in the alternation. The index of concordance is 0.7796408, which is just below the threshold of good performance (i.e.  $C \geq 0.8$ , Tagliamonte/Baayen 2012: 156).

<sup>12</sup> For the distributions of gerundial and infinitival clauses complementing *suggest* in the different varieties, see the Appendix (Table a).

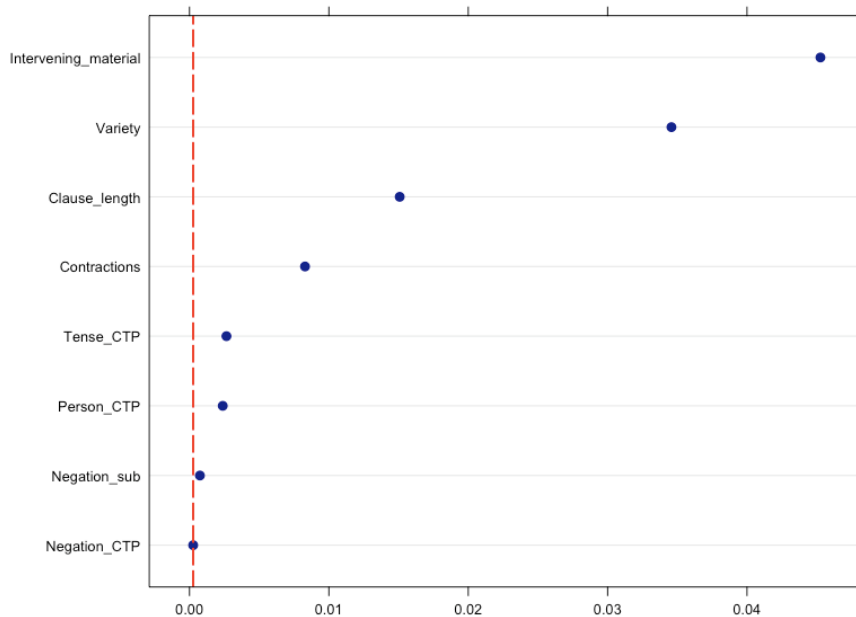


Figure 4: Random forest analysis for *suggest*.  $C = 0.7796408$ .

Figure 5 provides the logistic regression analysis for *suggest*. All levels are significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	0.184284	0.175610	1.049	0.293998	
VarietyBanE	0.719062	0.167569	4.291	1.78e-05	***
VarietyCanE	0.356119	0.155280	2.293	0.021825	*
VarietyJamE	-0.863874	0.194124	-4.450	8.58e-06	***
VarietyMaLE	0.409504	0.165986	2.467	0.013621	*
Person_CTP2nd	-1.147320	0.313950	-3.654	0.000258	***
Person_CTP3rd	-0.386626	0.123534	-3.130	0.001750	**
Negation_CTPYes	-0.838021	0.349285	-2.399	0.016429	*
Negation_subYes	-0.639835	0.281584	-2.272	0.023070	*
Tense_CTPPast	-0.275851	0.129966	-2.122	0.033796	*
Clause_length	-0.029435	0.005258	-5.598	2.17e-08	***
Intervening_material	0.161906	0.038474	4.208	2.57e-05	***
Contractions	0.228863	0.052389	4.369	1.25e-05	***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Figure 5: Logistic regression analysis for *suggest*.

The regression analysis indicates that JamE is the only variety that disfavours non-finite complementation more than BrE. The first person, more intervening material and more informal texts favour non-finite complementation. Negation, the past tense and longer CCs disfavour this complementation type.

The conditional inference tree (Figure 6) indicates that Variety is the most important predictive variable (Node 1) and that the varieties pattern differently. For JamE, there is relatively little internal variability in the alternation, as Clause Length is the only factor mentioned on the right side of the tree: longer CCs (more than 12 words) predict finite clauses more than shorter clauses in this variety (Nodes 16 and 17). For the other varieties, the tree mentions a wider range of predictive factors. If Clause Length is less than or equal to 16 words, Person of the CTP is a reliable factor (Node 3), with the first person predicting non-finite complementation more than the second and third person (Nodes 4 and 5). If Clause Length is higher than 16 words,

Formality is a predictive factor (Node 6). In the presence of contractions, negation in the subordinate clause has a higher predictive effect for finite complementation than subordinate clauses without negation (Nodes 13 and 14). In more formal texts (no contractions), Intervening material is a predictive factor (Node 7). In the context of intervening material, Variety comes into play again (Node 9): only BanE and MalE have a predictive effect for non-finite complementation when there is intervening material, whereas BrE and CanE predict finites in this context. This indicates that the favouring effect of the logistic regression analysis (Figure 5) for non-finites in the context of intervening material is restricted to BanE and MalE.

## 4 Discussion

Our hypothesis that there would be a diachronic rise in non-finite complementation in BrE has only been confirmed for *expect*. This means that *expect* has indeed become less iconic-isomorphic and more economic over time. In GloWbE, non-finite complementation is by far the preferred variant for all varieties. At the same time, the varieties (except for MalE) can be positioned along a cline of increasingly higher shares of non-finite complementation in accordance with their position in the Dynamic Model: BanE (phase 2) has the lowest share of non-finite complementation (88.9%), whereas CanE (phase 5) has the highest share (92.8%).<sup>13</sup> The CTP *expect* thus shows a second language acquisition effect in that the less advanced varieties have higher proportions of the more iconic variant (which is easier to acquire).

The CTP *suggest*, however, does not display a significant rise in non-finite complementation over time. Our hypothesis that an increasing share of non-finite complements in the various Englishes under study corresponds with an increasingly higher phase in Schneider's Dynamic Model is not confirmed either: the least advanced varieties – MalE and BanE – have the highest proportion of the non-iconic complementation pattern. The low frequency for non-finite complementation in JamE (19.9%) is also striking.

---

<sup>13</sup> Note that BrE, the reference variety, has a slightly higher percentage (94.3%). However, this is not significant ( $p = 0.198377442$ ).

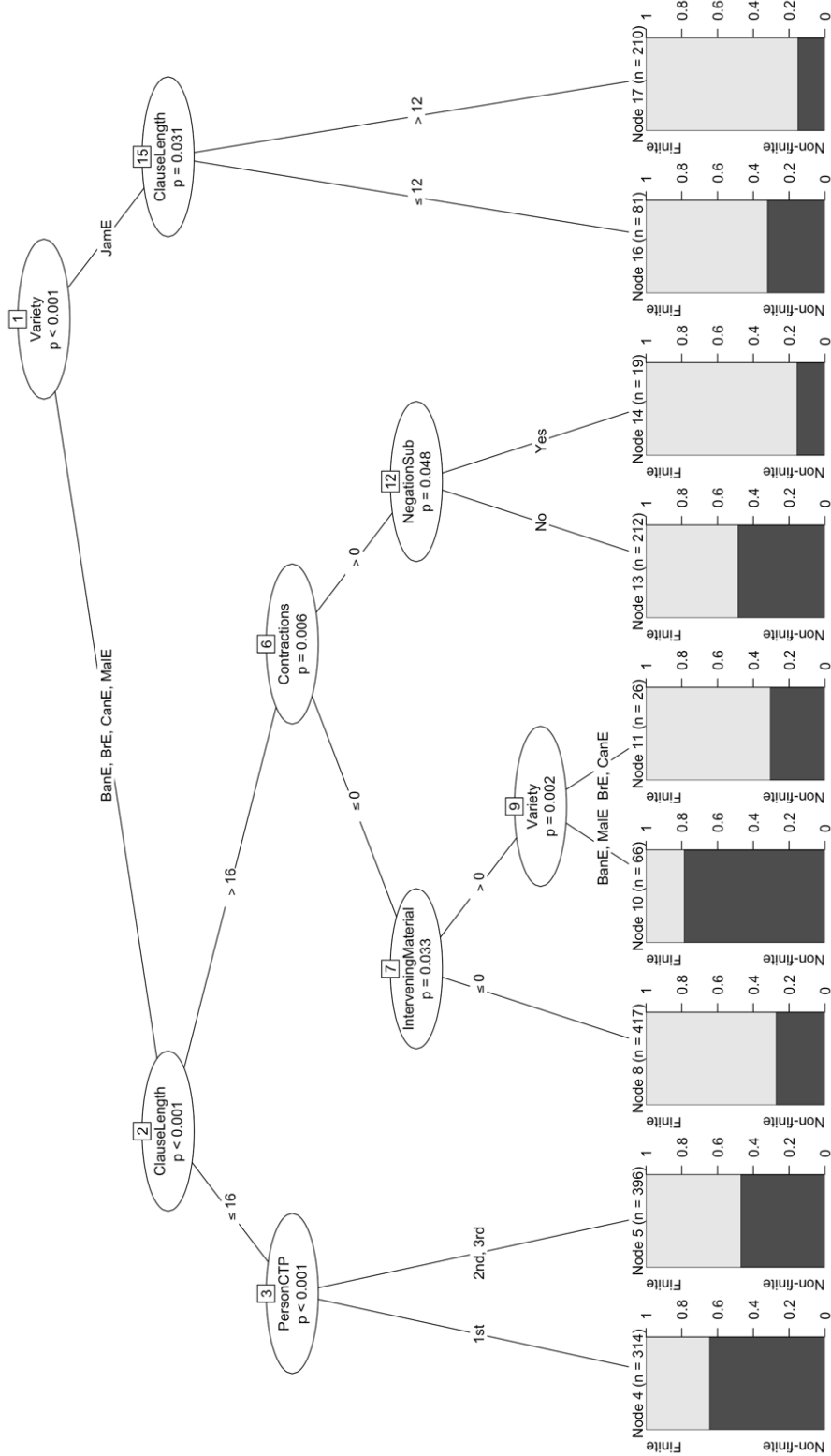


Figure 6: Conditional inference tree for *suggest*.

A possible explanation of these unexpected results for *suggest* and *expect* is substrate influence. In this respect, Patrick (2004: 423) states that Jamaican creole does not use “gerund forms with *-in(g)*”. This could explain why the gerund is so infrequent for *suggest* in the phase 4 variety. The high number of gerundial and infinitival clauses in BanE for *suggest* could be explained by the rich non-finite system in Bangla, which includes both an infinitive and a “*gerund-participle*” with “both nominal and verbal properties” (David/Connors/Chacón 2015: 229, 235). Finally, the high number of infinitival clauses in MalE for *expect* could be a transfer from Malay, which “does not have a finite/non-finite distinction. The base form of the Malay verb, which is used extensively especially in spoken/informal language, is [...] unmarked for tense” (Svalberg/Chuchu 1998: 33). However, more research is necessary to rule out other possible factors.

The statistical analyses in this article have indicated which factors play a role in the alternation between finite and non-finite alternation, which effect they have and how they interact. The random forest analyses give an overview of the most important factors in the decision between finite and non-finite complementation: for *expect*, Coreferentiality, Variety and Passive CTPs are the most important factors; for *suggest*, these are Intervening Material, Variety and Clause Length.

An overview of the hypotheses and results of the logistic regression analyses is given in Table 5.

Table 5: Hypotheses and results from the logistic regression analyses.

Factor	Hypothesis: preference for which complementation type?	Confirmed for <i>expect</i> ?	Confirmed for <i>suggest</i> ?
(More) intervening material	Finite	Not significant	No
Longer CCs	Finite	Yes	Yes
Negation CTP	Finite	Not significant	Yes
Negation subordinate clause	Finite	Yes	Yes
Passive CTPs	Finite	No	
Coreferentiality	Non-finite	Yes	
Formality	Finite	Yes (borderline significance)	Yes
Person CTP: 3rd	Finite	Not significant	Yes
Tense CTP: past	Finite	Not significant	Yes
Variety: BrE	Non-finite	Yes Exception: MalE	No



Table 5 indicates that the only factors related to the Complexity Principle that do not predict the hypothesized results are Intervening material (for *suggest*) and Passive CTPs (for *expect*): they favour non-finite complementation even though they constitute cognitively complex environments. However, two of the factors that measure cognitive complexity are not significant for *expect*. For *suggest*, the third person and the past tense significantly favour finite complementation.

However, the conditional inference trees indicate that there are differences between the varieties and between the two verbs concerning the predictive patterns. An important example is the factor Intervening Material: this only predicts non-finite complementation for the varieties MalE and BanE in the case of *suggest*. For CanE and BrE, more intervening material predicts finite complementation. This means that the favouring effect in the regression analysis is mostly influenced by the MalE and BanE data. For *expect*, however, the tree indicates that in MalE, intervening material predicts finite complementation slightly more than non-finite complementation. The conditional inference tree for *expect* also indicates that Variety is only a reliable factor for longer complement clauses without negation whose subject is not coreferential with the main clause subject.

## 5 Conclusion

In this article we have looked at the alternation between finite and non-finite complementation from a diachronic and a synchronic perspective. More specifically, we have investigated whether the observed rise of non-finite complementation in the history of BrE is reflected by distributional differences across World Englishes, whereby higher shares of non-finite CCs would be attested in varieties at a higher stage in Schneider's Dynamic Model. Underlying this research question is the hypothesis that just like clausal complementation may become less transparent over time, World Englishes will increasingly use less transparent complementation patterns as they move along the Dynamic Model. This is linked to a learner effect: less transparent complementation patterns are more difficult to acquire.

Our hypothesis is only borne out for *expect*, which shows a significant rise in non-finite complementation. Complementation patterns with *expect* thus become less iconic and more economic over time. It also shows a second language acquisition effect since, with the exception of MalE, the frequency of non-finite clauses in the varieties under study correlates with their level of advancedness in the Dynamic Model. For *suggest*, however, there is neither a significant rise of non-finite complementation over time nor a second language acquisition effect.

The random forest analyses indicate that for *suggest*, the most influential factors are Intervening material, Variety and Clause length; for *expect*, these are Coreferentiality, Variety and Passive CTPs. For both verbs, Variety thus appears to play a significant role.

The logistic regression analysis confirms most of our hypotheses. Non-coreferentiality, the third person, the past tense, formal texts, negation and longer CCs all favour finite complementation. Contrary to our hypotheses, however, the passive voice (*expect*) and more intervening material (*suggest*) favour non-finite complementation despite their status as cognitively complex environments. However, the conditional inference tree for *suggest* indicates that the latter favouring effect is variety-specific: only MalE and BanE favour non-finite complementation when there is intervening material following the CTP *suggest*.

Future research should examine our unexpected results in more detail, taking into account possible substrate influence and influence from prescriptivism, or could also include zero-*that*-clauses.

## Corpora

CLMET 3.0. A corpus of Late Modern English Texts. See [https://perswww.kuleuven.be/~u0044428/clmet3\\_0.htm](https://perswww.kuleuven.be/~u0044428/clmet3_0.htm)

BNC. British National Corpus. See <http://www.natcorp.ox.ac.uk>

GloWbE. Corpus of Global Web-Based English. See <https://corpus.byu.edu/glowbe/>.

## References

- Carter, Ronald et al. 2016. Suggest. *Cambridge Dictionary Online (English Grammar Today. An A-Z of Spoken and Written Grammar)*. Cambridge: Cambridge University Press. <https://dictionary.cambridge.org/grammar/british-grammar/common-verbs/suggest> [accessed 2018-6-5].
- Cuyckens, Hubert/ D’Hoedt Frauke. 2015. Variability in Clausal Verb Complementation: the Case of *Admit*. In: Höglund, Mikko et al. (eds.), *Perspectives on Complementation: Structure, Variation and Boundaries*, 77–100. London: Palgrave Macmillan.
- Cuyckens, Hubert/ D’hoedt, Frauke/ Szmrecsanyi, Benedikt. 2014. Variability in verb complementation in Late Modern English: finite vs. non-finite patterns. In: Hundt, Marianne (ed.), *Late Modern English Syntax*, 182–203. Cambridge: Cambridge University Press.
- David, Anne B./ Connors, Thomas J./ Chaçon, Dustin A. (eds.). 2015. *Descriptive grammar of Bangla* (Mouton-CASL Grammar Series 2). Berlin/Boston: De Gruyter Mouton.
- Denison, David. 1998. Syntax. In: Romaine, Suzanne (ed.), *The Cambridge History of the English Language. 1776-1997*, Vol. IV, 92-329. Cambridge: Cambridge University Press.
- Hawkins, John A. 1990. A Parsing Theory of Word Order Universals. *Linguistic Inquiry* 21(2). 223-261.
- Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Los, Bettelou. 2005. *The Rise of the To-Infinitive*. Oxford: Oxford University Press.
- Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
- Noonan, Michael. 2007 [1985]. Complementation. In: Shopen, Timothy (ed.), *Language Typology and Syntactic Description. Volume II: Complex Constructions*, Vol. 2, 2<sup>nd</sup> edn., 52-150. Cambridge: Cambridge University Press.
- Patrick, Peter L. 2004. Jamaican Creole: morphology and syntax. In: Kortmann, Bernd/ Schneider, Edgar/ Burridge, Kate/ Mesthrie, Rajend/ Upton, Clive (eds.), *A Handbook of Varieties of English. Morphology & Syntax: A Multimedia Reference Tool*, Vol. 2, 407-438. Berlin: Mouton de Gruyter.
- Quirk, Randolph et al. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Rohdenburg, Günter. 1995. On the Replacement of Finite Complement Clauses by Infinitives in English. *English Studies: A Journal of English Language and Literature* 76(4). 367-388.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7(2). 149–182.

- Rohdenburg, Günter. 2015. The Embedded Negation Constraint and the Choice between More or Less Explicit Clausal Structures in English. In: Höglund, Mikko et al. (eds.), *Perspectives on Complementation: Structure, Variation and Boundaries*, 101–127. London: Palgrave Macmillan.
- Schneider, Edgar. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
- Speelman, Dirk. 2014. Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In: Glynn, Dylan/ Robinson, Justyna A. (eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, 487–533. Amsterdam: John Benjamins Publishing Company.
- Steger, Maria/ Schneider, Edgar. 2012. Complexity as a function of iconicity: The case of complement clause constructions in New Englishes. In: Kortmann, Bernd/ Szmrecsanyi, Benedikt (eds.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*, 156-191. Berlin/Boston: De Gruyter.
- Svalberg, Agneta M-L./ Chuchu, Hjh Fatimah B. H. A. 1998. Are English and Malay worlds apart? Typological distance and the learning of tense and aspect concepts. *International Journal of Applied Linguistics* 8(1). 27–60.
- Szmrecsanyi, Benedikt et al. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109-137.
- Tagliamonte, Sali A./ Baayen, R. Harald. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135-178.
- Tizón-Couto, Beatriz. 2014. *Clausal Complements in Native and Learner Spoken English: A Corpus-Based Study with LINDSEI and VICOLSE*. Bern: Peter Lang AG.
- Van Driessche, Laetitia. 2018. *Clausal Verb Complementation in Varieties of English: A Corpus-Driven Approach*. Leuven: KU Leuven MA thesis.

## Appendix

Table a: Distributions for the clause types complementing *suggest* for the varieties in GloWbE. Normalised per 1,000,000 words.

Variety	Clause type	Raw frequencies	Normalised frequencies
British English	That-clause	203	13.6
	Gerundial clause	103	6.9
	Infinitival clause	18	1.2
Canadian English	That-clause	244	14.5
	Gerundial clause	184	10.9
	Infinitival clause	21	1.2
Jamaican English	That-clause	233	18.0
	Gerundial clause	43	3.3
	Infinitival clause	15	1.2
Malaysian English	That-clause	170	12.8
	Gerundial clause	72	5.4
	Infinitival clause	98	7.4
Bangladeshi English	That-clause	163	12.1
	Gerundial clause	95	7.0
	Infinitival clause	79	5.8