

Uncovering Variation in Word Englishes: Finite vs Non-finite Complementation of REMEMBER with Prospective Meaning

Cristina SUÁREZ-GÓMEZ & Raquel P. ROMASANTA

Author:

Cristina Suárez-Gómez
Universitat de les Illes Balears, Spain
cristina.suarez@uib.es
<https://orcid.org/0000-0002-5736-3705>

Raquel P. Romasanta
Universitat de les Illes Balears, Spain
raquel.pereira-romasanta@uib.es
<https://orcid.org/0000-0001-7508-4745>

Date of reception: 16/03/2023
Date of acceptance: 16/06/2023

Citation:

Suárez-Gómez, Cristina and Raquel P. Romasanta.
2023. "Uncovering Variation in Word Englishes:
Finite vs Non-finite Complementation of REMEMBER
with Prospective Meaning." ALICANTE JOURNAL OF
ENGLISH STUDIES 39: 35-54.
<https://doi.org/10.14198/raei.2023.39.02>

© 2023 Cristina Suárez-Gómez & Raquel P. Romasanta

License: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0):
<https://creativecommons.org/licenses/by/4.0/>



REMEMBER is a polysemous verb that can govern finite and non-finite complement clauses (CCs). This paper explores the variability between finite and non-finite CCs that follow REMEMBER with the prospective meaning ‘remember to do’ (as in *Remember to do your homework*), looking at three Asian World Englishes – Indian English, Sri Lankan English and Bangladeshi English – and British English, as represented in the *Corpus of Global Web-based English* (GloWbE). The prospective use of REMEMBER is traditionally classified as allowing *to*-infinitival CCs as the only complementation option (*Cambridge Dictionary Online*, “remember”; FrameNet; *Oxford Dictionaries Online*, “remember”; Huddleston and Pullum et al. 2002, 1242; Mair 2006, 215). However, large databases such as GloWbE reveal the existence of finite CCs with this meaning that depend on REMEMBER. The analysis of these CCs in competition, in terms of both distribution and a series of language-internal and external variables, confirms that those which increase complexity (e.g. longer CCs in number of words) favour the choice of finite CCs, in line with the Complexity Principle.

Keywords: REMEMBER; clausal complementation; (non-) finiteness; Complexity Principle; World Englishes; GloWbE

1. Introduction¹

REMEMBER is a retrospective verb with three meanings: ‘recall’, ‘bear in mind the fact’ and ‘remember to do’ (Fanego 1996; Mair 2006; Cuyckens, D’hoedt & Szmrecsanyi 2014), illustrated in (1) to (3) below. In terms of clausal complementation, REMEMBER can govern both finite and non-finite complement clauses (CCs), italicized in the examples. However, these CCs are not always freely interchangeable, but rather are determined by the meaning of the verb. For example, the meaning ‘bear in mind the fact’ allows only declarative CCs (Mair 2006, 226), both expanded (introduced by the complementizer *that*) and bare (without complementizer), as shown in examples (4) and (5).

(1) I remember *we were analysing a rather harmless poem about picking blackberries*. (GloWbE Great Britain (GB))

(2) Remember *you’re at a party*. (GloWbE Bangladesh (BD))²

(3) I must remember *to take a carrier bag and to park a bit closer!* (GloWbE GB)

(4) You must remember *that this is a secret*. [*Oxford Dictionaries Online*, “remember”]

(5) You must remember *∅ this is a secret*.

Most studies of this verb have focused on the variation between finite and non-finite CCs that typically occurs with the retrospective meaning ‘recall’, illustrated in (6) and (7) respectively (see, for example, Fanego 1996; Mair 2006; Cuyckens, D’hoedt & Szmrecsanyi 2014).

(6) I do remember *that I had seen a permanent link to the images attached with this post before I went offline*. (GloWbE BD)

(7) I do remember *seeing a permanent link to the images attached with this post before I went offline*.

1 We are deeply grateful to Lucía Loureiro and the two reviewers and the editors of RAEI for their insightful comments, which have improved enormously. We are also grateful to the generous financial support of grant PID2020-117030GB-I00, funded by MCIN/AEI/10.13039/501100011033.

2 Examples drawn from GloWbE are reported as they appear in the original corpus, even

The distribution of CCs after REMEMBER with the prospective meaning ‘remember to do’, another of the most frequently-attested meanings of the verb, is often neglected, since, according to various grammars, dictionaries and databases (*Cambridge Dictionary Online*, “remember”; FrameNet; *Oxford Dictionaries Online*, “remember”), it only allows *to*-infinitival CCs as a complementation option, as illustrated in (8). Therefore, it is not expected to show variability in its complementation profile.

(8) Please remember *not to fill this empty space with some useless but “pretty” design elements.*

However, an analysis of a sample of REMEMBER in a large corpus, *the Corpus of Global Web-based English* (GloWbE, Davies 2013), reveals that variation between finite and non-finite CCs after REMEMBER exists not only when it means ‘recall’, but also when it means ‘remember to do’, as illustrated in (9), which contrasts with (8) above. This reveals a new envelope of variation for this verb.

(9) Please remember *that you shouldn’t fill this empty space with some useless but “pretty” design elements.* (GloWbE India (IN))

Both (8) and (9) conform to the definitions provided for the prospective meaning ‘remember to do’. For example, “do something that one has undertaken to do or that is necessary or advisable” (*Oxford Dictionaries Online*, ‘remember’) and “[a] cognizer thinks of and performs an action that is a self- or other-imposed task or some other kind of desirable behavior” (FrameNet). In (8), the *to*-infinitival CC conveys the nuance that the task of *filling this empty space* is self- or other-imposed, and/or that it is a desirable behaviour, in the same way as in (9), where this (self- or other) imposition to *fill this empty space* is conveyed by the modal verb *should*.

The relevance of modal verbs is also reflected in the modal *should* and the semi-modal *have to* in (10) and (11): here the finite declarative CCs depend on REMEMBER with the meaning ‘remember to do’, since they describe some kind of behaviour or task which is either self-imposed or other-imposed and which is necessary or advisable to perform; that is, they convey deontic obligation. In these cases, the lexical verb in the CC remains non-finite.

(10) Always remember though, *to ensure good diversification you should always ensure you spread your money over at least 200 loans.* (GloWbE GB)

(11) Remember *that you have to get to that interview!* (GloWbE IN)

Crucially, finite declarative CCs may have a non-finite CC counterpart in the form of a simple *to*-infinitival CC, as shown in examples (12) and (13). Another key element relating to the fact that the finite CCs in (10) and (11) are interchangeable with the non-finite CCs included in (12) and (13) is that the subject of the main clause (MC) and the CC is the same, the second person pronoun *you*.

(12) Always remember though, *to spread your money over at least 200 loans to ensure good diversification*.

(13) Remember *to get to that interview*.

The aim of the present study, then, is to examine this unexplored envelope of variation between finite and non-finite CCs after *REMEMBER* with the meaning ‘remember to do’. First, from a quantitative perspective, examining the distribution of finite and non-finite CCs across different varieties of English. And second, from a quantitative and qualitative perspective, through a survey of a series of language-internal variables (e.g. meaning of the CC verb and complexity of the CC in number of words, see Section 3.1 below) which might determine the choice of one CC-type over the other and thus explain when finite CCs are an alternative to non-finite CCs.

The article is structured as follows. Section 2 offers an overview of the clausal complementation profile of the verb *REMEMBER* and reviews those studies which have explored CC-variation after this verb. Section 3 describes the methodology and the variables used in the analysis, after which Section 4 discusses the results of the analysis. Finally, Section 5 summarises the main conclusions and the implications of the study.

2. Background

In Present-day English, the polysemous verb *REMEMBER* is classified as a catenative verb that belongs to the retrospective class of verbs, together with others such as *REGRET* and *FORGET*. With this set of retrospective verbs, non-finite CCs are considered not to be freely interchangeable, with *to*-infinitive clauses expressing a prospective meaning or *-ing* clauses with a retrospective one (Fanego 1996, 71; Huddleston & Pullum 2002, 1242).

Despite formulating slightly different definitions of the meanings of the verb *REMEMBER*, dictionaries and databases such as FrameNet agree on the fact that this verb takes clausal complements with three different meanings: (i) ‘recall’; (ii) ‘bear in mind the fact’ and (iii) ‘remember to do’.

Most studies of *REMEMBER* deal with the meaning ‘recall’, mainly due to the variation between finite and non-finite CCs in its complementation profile (cf. Fanego 1996; Mair 2006; Cuyckens, D’hoedt & Szmrecsanyi 2014). Finite

CCs are declarative CCs, both when expanded with the complementizer *that* and when bare (without a complementizer), as in (14) and (15) respectively. Finite CCs have been available after REMEMBER since Middle English (Fanego 1996, 74). Concerning non-finite CCs, the perfective *to*-infinitival CC alternative (see example (16)) was historically a frequent option. Nowadays, however, the most common non-finite option is a gerund-participial CC, illustrated in (17). Gerund-participial CCs emerged as a new complementation option for REMEMBER during the Late Modern English Period (De Smet 2013), replacing other options.

(14) I vividly remember *that as a child I was awfully thin*. (GloWbE Sri Lanka (LK))

(15) I vividly remember \emptyset *as a child I was awfully thin*.

(16) [...] any one of them remembered *to have received some help from the poor Brahmin*. (GloWbE LK)

(17) I do not remember *hearing any of the Anglo-Saxon words even through four years at Oxford University*. (GloWbE BD)

Research into REMEMBER with this meaning has analysed CC-variability in English from Late Modern English onwards. Cuyckens, D'hoedt & Szmrecsanyi (2014) address the earlier period and García-Castro (2020) focuses on Present-day British English (BrE) and Outer Circle varieties of English such as Indian, Bangladeshi and Sri Lankan Englishes. Both studies explore the distribution of finite and non-finite CCs after REMEMBER and use binary logistic regression analysis to determine the factors that favour each variant. Cuyckens, D'hoedt & Szmrecsanyi (2014) also include DENY and REGRET in their study.

Both studies select largely semantic and structural factors (cf. Section 3.1 for more details on these variables). The structural factors –nine out of a total of fourteen– are closely related to the Complexity Principle (Rohdenburg 1996, 2006), whereby more cognitively complex syntactic environments (e.g. negative structures, passive structures, long sentences, among other features) would favour the use of more explicit grammatical options (Rohdenburg 1996, 51). In CC-variation, a difference exists in the grammatical explicitness of the clauses (Rohdenburg 1996, 151-152). While finite declarative CCs are grammatically more explicit, in that they have an expressed subject and they code for tense/mood distinctions, for instance (Cuyckens, D'hoedt & Szmrecsanyi 2014, 199), non-finite CCs are the less grammatically explicit option. With this in mind, it is expected that structurally complex environments (structural complexity being a measure of cognitive complexity) would favour the more explicit option, that is, finite CCs. Conversely, less complex environments would favour the non-

finite alternative (Cuyckens, D'hoedt & Szmrecsanyi 2014, 198-199). Mair (2006) supports this idea in his analysis of the distribution of CCs after REMEMBER meaning 'recall'; although the variant selection remains free, finite CCs are more likely to occur in contexts of "a high degree of elaboration" (2006, 222), as in examples (18) and (19), which he describes as "clumsy stylistically or downright unacceptable" (Mair 2006, 222).

(18) The readers will remember *that in the heyday of non-co-operation, the terms 'Mr.' and 'Esquire', were dropped by Congressmen and the nationalist press, and 'Shri' was the title largely used for all, irrespective of religion.* (1938 M. K. Gandhi in D. G. Tendulkar Mahatma (1952) IV. 348; cf. Mair 2006, 223)

(19) ? readers will remember *the terms 'Mr.' and 'Esquire' being dropped by Congressmen and the nationalist press, and 'Shri' was the title largely used for all, irrespective of religion, in the heyday of non-co-operation* (Mair 2006, 223)

Cuyckens, D'hoedt & Szmrecsanyi (2014) find that non-finite complementation increases from the first subperiod of the Late Modern English Period (1710-1780) to the second subperiod (1781-1920). As to the factors favouring each of the variants, they conclude that semantic and structural factors, as well as some additional ones, condition CC-variation. Regarding structural variables related to the Complexity Principle (Rohdenburg 1996, 2006), Cuyckens, D'hoedt & Szmrecsanyi's results are conflicting. There are features considered to increase cognitive complexity which disfavour non-finite complementation (e.g. different subject between the MC and the CC, as in (20)), but others (e.g. complexity of the CC) that favour them. They conclude that "Rohdenburg's proposed disfavouring effect cannot be generalised to all structural complexity factors, and in that sense, Rohdenburg's Complexity Principle does not apply as generally as commonly held" (2014, 199).

(20) While driving a car or any vehicle for that matter, one must *remember that you need to exercise caution.* (GloWbE IN)

In a similar analysis of Present-day British, Indian, Bangladeshi, and Sri Lankan Englishes, García-Castro (2020) finds that non-finite complementation is more common in all the varieties of English surveyed, but that the tendency is less pronounced in L2 varieties. García-Castro argues that this is related to the interplay of factors that condition the evolution of these L2 varieties (Schneider 2007, 88-90, 99-107), among which innovation, language contact and second-language acquisition (SLA) are notable.

Within the broad phenomenon of innovation, we find processes which are said to lead to the simplification of morphosyntactic features across L2 varieties

of English developed around the world (Williams 1987). One of these processes is the tendency towards transparency. Based on such a tendency, finite CCs should be favoured by L2 speakers because they are more transparent in that, unlike non-finite CCs, they provide more conceptual information (meaning) in a more explicit manner (expressed subjects, coding for tense and mood, among others) (Steger & Schneider 2012, 156-157).³

Together with innovation, understood here in terms of transparency, the other two factors mentioned above that usually intersect –language contact and SLA processes– are also considered potential factors which may condition the selection of a finite or a non-finite CC (Schneider 2013, 143, 148; cf. Thomason 2001). García-Castro (2020) surveys the complementation profile of the five most spoken languages across India, Bangladesh and Sri Lanka (Hindi, Telugu, Sinhala, Tamil and Bengali; cf. Eberhard, Simons & Fennig 2019) and finds that all these substrate languages have a structure equivalent to declarative CCs (cf. Annamali & Steever 1998; Krishnamurti 1998; Steever 1998; Kachru 2006; Dasgupta 2007; Gair 2007; Chandralal 2010; Thompson 2012). This may explain why declarative CCs are more frequent in L2 varieties than in BrE: L2 speakers are said to select the structures closest to those available in their L1(s), in line with the ‘Shortest Path Principle’ and the ‘Teddy Bear Principle’. The former states that, when there are rules allowing for variation between more than one option in the target language, learners will tend to select only one, usually the one closer to the corresponding L1 feature (Wald 1996). Meanwhile, the ‘Teddy Bear Principle’ posits that learners show a tendency to stick to structures which they know well and can use safely (Hasselgren 1994; Biewer 2015, 14). All these factors potentially account for the divergence in the proportion of non-finite CCs between British English and the L2 varieties (García-Castro 2020).⁴

By contrast, the meanings ‘remember to do’ and ‘bear in mind the fact’ have received less attention. One of the reasons for this is that their complementation profile has, to date, exhibited no variation between finite and non-finite CCs (Mair 2006). In the case of the meaning ‘remember to do’, since REMEMBER entered the English language, only simple *to*-infinitival CCs, as in (21), have been recorded as a complementation option after this meaning.⁵ Unlike the previous

3 One reviewer suggests that the fact that finite CCs are more often used in L2s may be because they are lagging behind the L1, which are increasingly using non-finite complements. Although this is a very interesting hypothesis, we cannot confirm it because there are no samples of comparable historical data.

4 See García-Castro (2020) for further information on the factor weights that determine variation in the divergence between finite and non-finite clauses depending on REMEMBER ‘recall’. In line with Cuyckens, D’hoedt & Szmrecsanyi (2014), both semantic (e.g. meaning CC) and structural factors (e.g. complexity CC in nr of constituents) prove to be determinant in this context.

5 By Middle English, *to*-infinitival CCs were already available after REMEMBER with a prospective meaning (Cambridge Dictionary Online, “remember”; OED Online, “remember”; Oxford

two meanings, ‘remember to do’ has a prospective orientation, since the action referred to by the complement clause is projected into the future, and this is what makes the *to*-infinitive a felicitous complement here, since *to*-infinitives have prospective origins. In (21), in order to do the shopping, first you need to remember that you do indeed have to perform this action.

(21) Did you remember *to do the shopping*? (*Cambridge Dictionary Online*, s.v. “remember”)

However, as we noted above, large databases of contemporary Englishes around the world such as GloWbE show cases of finite CCs after REMEMBER coexisting with non-finite ones when the meaning of the verb is ‘remember to do’. This demonstrates that while complementation patterns may show broad trends, with verbs with similar meanings taking similar complements (e.g. memory verb FORGET, Huddleston & Pullum 2002, 1242), the question of whether any individual verb will appear with any particular complement depends not only on its meaning but also on usage and entrenchment, which in turn depend on frequency. The coexistence of finite and non-finite CCs with REMEMBER with other meanings, as well as with other related verbs (e.g. FORGET), may have caused the increase of variants for REMEMBER with the meaning ‘remember to do’.

In the current study, we replicate the same kind of analysis of REMEMBER with the meaning ‘recall’ conducted by Cuyckens, D’hoedt & Szmrecsanyi (2014) and García-Castro (2020) for the meaning ‘remember to do’. In this context, there are also finite and non-finite CCs in competition (see Table 1 in Section 4) and thus the Complexity Principle can be considered a potential factor influencing CC-variation. Moreover, since the data analysed relate to four varieties of English, one L1 from the Inner Circle in Kachru’s model (British English; Kachru 1982, 356) and three L2 varieties from the Outer Circle (Indian, Bangladeshi, and Sri Lankan Englishes), those factors conditioning the evolution of these L2 varieties found to be relevant in García-Castro (2020) (innovation, language contact, and SLA processes) may also influence CC-distribution after REMEMBER with this meaning.

3. Data and Methodology

3.1. Corpus

The data used to explore the envelope of variation shown by REMEMBER with the prospective meaning ‘remember to do’ were extracted from the *Corpus of Global Web-based English* (GloWbE, Davies 2013), specifically from the British,

Dictionaries Online, “remember”; Visser 1963-1973, 2234; Fanego 1996, 75).

Indian, Bangladeshi and Sri Lankan sections. The data in GloWbE have been described as a sample of “English as used on the web” (Loureiro-Porto 2017, 455). Since the data have been retrieved recently from the Internet, they provide the opportunity to explore synchronic variation across varieties English. Another advantage of this corpus is its size: 1.9 billion words. Unlike corpora such as ICE (*The International Corpus of English*), which proves to be too small for the present analysis (García-Castro 2018), low-frequency phenomena such as CC-variation can be explored successfully using GloWbE.

In order to compile the data set, we retrieved from the online interface all the hits available for the forms <remember>, <remembers>, <remembered>, and <remembering> in four sections of GloWbE (GloWbE GB, GloWbE IN, GloWbE BD and GloWbE LK). We then randomised the sample in an Excel spreadsheet and selected the first 3,000 hits from each section, obtaining a sample of 12,000 examples. Data were manually pruned of all spurious (e.g. repeated examples) and irrelevant (e.g. nominal complements) hits, leaving the total number of valid examples at 4,931. Then, we classified all instances of REMEMBER + CC based on the meaning ‘recall’ (1,508 examples), ‘bear in mind the fact’ (2,897 examples) and ‘remember to do’ (526 examples).

3.2 Objectives

The aim of the present study, as mentioned in the introduction, is to examine the unexplored envelope of variation between finite and non-finite CCs after REMEMBER with the meaning ‘remember to do’. The objectives are twofold:

- (i) First, explore the frequency distribution of the two complementation patterns (finite and non-finite) in competition from a quantitative perspective across the four English varieties.
- (ii) Second, through a quantitative and qualitative analysis of a series of language-internal variables (e.g. meaning of the CC verb and complexity of the CC in number of words, see Section 3.2.1 below), determine the factors that condition the choice of one CC-type over the other and thus explain when finite CCs are an alternative to non-finite CCs.

3.2 Variable Analysis

3.2.1 Coding the Variables

Each corpus attestation of REMEMBER + CC with the meaning ‘remember to do’ was coded for a number of intra-linguistic factors which might condition CC-variation. These are characteristics of the MC, characteristics of the CC, and

characteristics of the combined structure of the MC and the CC. The selection of potentially significant factors, listed in Table 1, was drawn from those in Cuyckens, D'hoedt & Szmrecsanyi (2014)⁶ and García-Castro (2020).

TABLE 1. List of language-internal variables coded

Type of variable	Name of variable	Values
Semantic factor ⁷	MEANING CC	Event/Action State
Structural factors ⁸	SUBJECT	Pronoun Non-pronominal Non-expressed
	COMPLEX CC - CONSTITUENTS	Short (0-1) Medium (2-3) Long (4-5)
	COMPLEX CC - WORDS	Short (1-7) Medium (8-14) Long (+15)
	INTERVENING MATERIAL	Presence Absence
	SUPPLEMENTATION	Presence Absence
	COORDINATION	Presence Absence
	POLARITY CC	Positive Negative

6 Factors such as MEDIUM and PERIOD were not included in the present study because they do not apply to our data: we treat GloWbE as representing English as used on the web (no variation regarding MEDIUM) and the data is from the same period, 2012, and therefore there will not be any variation regarding period.

7 Within the semantic factors, we also checked the TIME REFERENCE between the MC and the CC, which can either be “independent TIME REFERENCE (ITR)” or “dependent TIME REFERENCE (DTR)”. However, these were excluded from the final analysis because they show collinearity, in that the *to*-infinitive CCs are by default DTRs (e.g. *one must remember that you need to exercise caution.* (GloWbE IN)) and the finite declarative CCs are ITRs (e.g. *Remember to convince yourself and add as much belief as possible without unnecessary stress being placed upon yourself.* (GloWbE GB)).

8 Within the structural factors, VOICE OF THE CC VERB was initially included, but discarded from the final analysis because there is only one instance of the verb in the passive voice in the whole database, and thus this factor behaves as a categorical one.

As shown in Table 1, MEANING CC is the semantic factor considered, with the same values used by Cuyckens, D’hoedt & Szmrecsanyi (2014) and García-Castro (2020). These values are *event/action*, as in (25) and *state*, as in (26).

(25) You must remember *to look at train timetables, hostels to stay at and make a note of the must-see sightseeing sights as your time in each country is limited (if you wish to take advantage and visit as many countries as possible within your time-span)*. (GloWbE GB)

(26) I think people need to remember *to be patient*. (GloWbE IN)

Structural factors represent the majority of the variables analysed, as illustrated in Table 1. As noted above (Section 1), one of the features in the alternation of finite and non-finite CCs with prospective REMEMBER is that the subject of the MC and the CC are always coreferential. Therefore, the variable SUBJECT has three variants: pronominal, non-pronominal and non-expressed subject. For this variable, in the case of non-expressed subject, we examined the “understood subject”. In example (25) above, the understood subject is *you*, in agreement with the explicit subject of the main clause. The complexity of the CC is measured based on its predicate and in two different ways: the number of constituents, and the number of words. Including both measures allows for a more comprehensive analysis, since a single constituent may contain several words, and on the other hand several constituents may be expressed with very few words. Features that might increase the cognitive complexity of the CC are accounted for by means of the variables INTERVENING MATERIAL, SUPPLEMENTATION, COORDINATION, and POLARITY CC VERB. SUPPLEMENTATION refers to “elements which occupy a position in linear sequence without being integrated into the syntactic structure of the sentence” (Huddleston & Pullum 2002, 1350). This should also add complexity to an environment and, according to the Complexity Principle (Rohdenburg 1996, 2006) SUPPLEMENTATION prompts the use of finite CCs.

3.2.2. Statistical analysis

As Cuyckens, D’hoedt & Szmrecsanyi (2014) explain, a binary logistic regression analysis is a suitable way of analysing free variation between CCs. Binary logistic regression is a statistical test that predicts the choice between two variants (the values of the dependent variable) based on a series of explanatory factors (i.e. each of the variables presented in Section 4.2.1) (Field 2009, 265). However, as will be shown in Section 4.1, the low number of finite CCs does not allow us to conduct such a test, and thus Pearson’s chi-square test and Fisher’s exact test were carried out instead.

Pearson's chi-square test shows whether there is a relationship between two categorical variables (the type of variables we use) (Field 2009, 688). The chi-square test has two important assumptions:

- i. The assumption of the independence of the data.⁹
- ii. The assumption that the expected frequencies in each cell should be greater than five. In contingency tables larger than 2x2, it is acceptable to have up to 20% of the expected frequencies below five, although this results in a loss of statistical power. Expected frequencies should never be below one (Field 2009, 691-692).

When the second assumption of the chi-square test is violated, the test itself is no longer reliable. In the case of 2x2 contingency tables and small size samples, Fisher's exact test is usually applied. This test is "a way of computing the exact probability of the chi-square statistic" (Field 2009, 690) and is designed to overcome the problem of small samples. This is the case with the data for the envelope of variation analysed here, the results of which we discuss below.

4. Analysis and results

4.1. Overall Distribution

Table 2 shows the overall distribution of finite and non-finite CCs depending on REMEMBER meaning 'remember to do'.

TABLE 2. Distribution of finite and non-finite CCs with the meaning 'remember to do'

	No.	%
<i>Finite CCs</i>	36	7.0
<i>Non-finite CCs</i>	490	93.0

Non-finite CCs, represented by *to*-infinitival CCs, are the default variant, accounting for 93% of the examples, as opposed to finite declarative CCs, representing only 7% of the total hits.

⁹ Although more than one datapoint have been drawn from the same blog/website, this does not mean that they come from the same author taking into account that many of these webpages are blogs and more than one person is participating. Unfortunately, this cannot be checked because many of these sources are no longer operative.

In terms of variety, all sections of GloWbE replicate these global results, favouring non-finite CCs, as reflected in Table 3.

TABLE 3. Distribution of finite and non-finite CCs after REMEMBER with the meaning ‘remember to do’ in the four sections of GloWbE

	GloWbE GB		GloWbE IN		GloWbE BD		GloWbE LK		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Finite CCs	5	4.0	13	9.5	14	9.0	4	4.0	36	7.0
Non-finite CCs	117	96.0	125	90.5	144	91.0	104	96.0	490	93.5

It is important to note that the varieties with the highest proportions of finite CCs are two of the L2s (GloWbE IN and GloWbE BD). The result of the chi-squared tests, of all varieties together on the one hand and the individual L2 varieties compared to GB on the other, show that the different distributions are not statistically significant. However, the distributions seem to align with the alleged tendency of L2s to favour finite CCs to a greater extent than BrE (cf. Steger & Schneider 2012); the exception here is GloWbE LK, which presents a tendency similar to GloWbE GB.

Within declarative CCs, it is interesting to observe the distribution of different forms, that is, expanded declarative CCs introduced by the complementizer *that* and bare declarative (without complementizer) CCs. These results are set out in Table 4.

TABLE 4. Distribution of expanded and bare declarative CCs after REMEMBER with the meaning ‘remember to do’ in the four sections of GloWbE

	GloWbE GB		GloWbE IN		GloWbE BD		GloWbE LK		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Expanded declarative CCs	3	60.0	9	69.0	6	43.0	-	-	18	50.0
Bare declarative CCs	2	40.0	4	31.0	8	57.0	4	100	18	50.0
Total	5	100	13	100	14	100	4	100	36	100

Although the numbers are too low to reach definite conclusions, some tendencies can be identified. Expanded declarative CCs, which are more transparent structures due to the presence of the complementizer *that* (Steger & Schneider 2012), are hypothesised to be favoured in L2 varieties. This indeed can be seen in GloWbE IN, which scores higher in expanded declarative CCs than any of the other varieties. Regarding GloWbE BD and GloWbE LK, the respective substrate

languages may be interfering in the higher frequency of bare declarative CCs, since Bengali and Sinhala, the most widely spoken languages in Bangladesh and Sri Lanka (cf. Chandralal 2010; Thompson 2012; Eberhard, Simons & Fennig 2019), have structures equivalent not only to expanded declarative CCs but also to bare declarative CCs. Speakers of these varieties may consider bare declarative CCs to be closer to structures in their L1s (following the ‘Shortest Path Principle’) and feel as comfortable using them (following the ‘Teddy Bear Principle’) as they do in using expanded declarative CCs.

Of the eight variables included in Table 1 in the methodology section (see 3.2.1) only three proved to be statistically significant at $p < 0.05$:¹⁰ COMPLEX CC - WORDS (cf. Table 5 and Figure 1), and SUPPLEMENTATION and COORDINATION in GloWbE LK (cf. Table 6 and Table 7). The remaining variables, whether globally and/or across sections of GloWbE (MEANING CC, COMPLEX CC - CONSTITUENTS, INTERVENING MATERIAL, and POLARITY CC) were seen to be non-significant, mostly because the distribution between finite and non-finite CCs is unbalanced. For the variable SUBJECT the low number of finite CCs in this envelope of variation leads to a sparseness of data in the tables, and thus the application of statistical tests is ruled out.

The first significant variable, COMPLEX CC - WORDS, contains the three simplified values, *short*, *medium* and *long*, and all of these favour the use of the non-finite alternative, as can be seen in Table 5. However, it is important to note the positive correlation between the increase in the number of words of the CC and the larger number of finite CCs: only 2.0% of the *short* CCs are finite, 6.5% of the *medium* CCs are finite and, finally, 14.0% *long* CCs are finite, the largest percentage of finite CCs. These results are significant and confirm Rohdenburg’s Complexity Principle (1996, 2006) (cf. Section 2), according to which more complex environments should favour more explicit structures. Although the non-finite option is favoured in all contexts, Table 5 confirms the significant preference for finite CCs when there are more words in the CC. In fact, a larger number of words would make the environment more complex and justify the increase in the use of finite CCs, which are more explicit than non-finite CCs (Rohdenburg 2006; Steger & Schneider 2012). These results are significant at $p < 0.05$. Therefore, this variable does determine the finiteness of the CC.

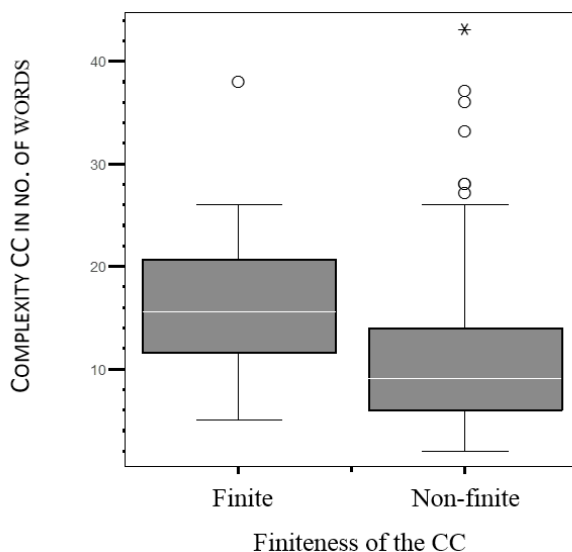
¹⁰ We applied Pearson’s chi-square (and Fisher’s exact test when the chi-square was not possible).

TABLE 5. Distribution of finite and non-finite CCs in variable
 COMPLEX CC - WORDS ($\chi^2 = 16.553, df = 2, p = 0.000$)

	Short (1-7)		Medium (8-14)		Long (+15)	
	No.	%	No.	%	No.	%
<i>Finite CCs</i>	4	2.0	14	6.5	18	14.0
<i>Non-finite CCs</i>	176	98.0	204	93.5	110	86.0

The fact that finite CCs are longer in number of words than non-finite CCs is reflected in the box-plot in figure 1, which uses the values of the variable as a scale ranging from one to 43 words. The left part of the graph represents finite CCs; the median is around 15-16 words, and the remaining 50% of instances, in the grey box, are between 12-21 words long with only one outlier in the range 35-40 words. On the other hand, non-finite CCs, which are represented on the right part of the graph, display more outliers, yet 50% of instances show that they are shorter than their finite counterpart. The graph shows that the median is around eight to nine words and that 50% of non-finite CCs range from five to six words to 14-15 words. In addition, the least extreme value is shorter in the case of non-finite CCs, represented by the bottom whisker, when compared to the finite one. This is in line with the results from Table 5 and also with Rohdenburg’s Complexity Principle.

FIGURE 1. Boxplot of the variable COMPLEX CC - WORDS



Turning to the results for the variables SUPPLEMENTATION (cf. Table 6) and COORDINATION (cf. Table 7), these are found to be significant in GloWbE LK. Examples (27) and (28) illustrate the presence of supplementation and coordination respectively. Supplementation refers to “elements which occupy a position in linear sequence without being integrated into the syntactic structure of the sentence” (Huddleston & Pullum 2002, 1350), as *especially since pale yellow tends to bleed into a white background* in (27). As for coordination, its presence means that several structures are coordinated after the verb REMEMBER, in the case of (28) two bare declarative CCs.

(27) Remember to use a dark color as contrast to yellow especially since pale yellow tends to bleed into a white background. (GloWbE LK)

(28) Remember you need to be passionate about your business, always have the right attitude, and be disciplined in completing your goals. (GloWbE BD)

TABLE 6. Distribution of finite and non-finite CCs in variable SUPPLEMENTATION in GloWbE LK (Fisher’s exact test = 0.01)

	Presence of supplementation		Absence of supplementation	
	No.	%	No.	%
<i>Finite CCs</i>	2	40.0	2	2.0
<i>Non-finite CCs</i>	3	60.0	101	98.0
Total	5	100	103	100

TABLE 7. Distribution of finite and non-finite CCs in variable COORDINATION in GloWbE LK (Fisher’s exact test = 0.0205)

	Presence of coordination		Absence of coordination	
	No.	%	No.	%
<i>Finite CCs</i>	2	28.5	2	2.0
<i>Non-finite CCs</i>	5	71.5	99	98.0
Total	7	100	101	100

Tables 6 and 7 show that for presence of supplementation and coordination, the percentage of finite CCs is greater than in the case of absence (40.0% vs 2.0% and 28.5% vs 2.0%), which seems to be in line with the Complexity Principle (1996, 2006). As this principle states, in a more cognitively complex environment,

created by the presence of supplementation (see Rohdenburg 1996, 2006) and coordination, the preference for finite CCs, that is, the more explicit option, increases.

5. Conclusion

This paper has identified a hitherto unexplored complementation pattern of the verb REMEMBER with the meaning ‘remember to do’. Together with *to*-infinitive CCs, finite declarative CCs are the default structure in this context. Thus, our study has shown that we cannot exclusively rely on information about complementation patterns from dictionaries and grammars, even corpus-based ones, since there are some patterns which might be so infrequent that they only emerge in a very large corpus; however, they still need to be researched.

We have proposed and explored a second envelope of variation after the verb REMEMBER, which is different from the well-known variation after REMEMBER meaning ‘recall’. Probabilistic variation between finite and non-finite CCs after REMEMBER also appears to take place when the meaning is ‘remember to do’, and we have considered this across four varieties of Present-day English. We have (i) explored the frequency distribution of the two patterns in competition, and (ii) conducted a variable analysis of the intra-linguistic factors that may condition the choice of CC here. Based on previous studies on the verb REMEMBER (Cuyckens, D’hoedt & Szmrecsanyi 2014; García-Castro 2020), we have included language-internal factors among these, features related to Rohdenburg’s Complexity Principle.

The study has shown that, as expected, non-finite CCs are clearly predominant in all the individual sections of GloWbE. Nevertheless, finite CCs have emerged as a possible alternative, since the meaning conveyed remains the same when a *to*-infinitival CC is transformed into a declarative CC, and vice versa.

The variable analysis carried out reveals that finite CCs reflect a pattern not previously explored in the literature. In this analysis three variables show significant results (at $p < 0.05$): COMPLEX CC - WORDS across the whole data set, and SUPPLEMENTATION and COORDINATION in the case of GloWbE LK. The results of the analysis of these variables point towards the Complexity Principle as a potential cause of the increased use of finite CCs against the overwhelming use of non-finite CCs with this meaning of REMEMBER. The significance of the same complexity factors previously noted in the literature points to the pervasive impact of production constraints, overriding other tendencies, such as the diachronic trend of the *to*-infinitive outcompeting the finite clause.

The particular relevance of this study is that it has revealed an unexplored envelope of variation and established the basis for future research here. Larger sets of data are necessary to assess whether the distribution identified across varieties

of English in the present study is significant and which factors are responsible for it, especially those that condition the evolution of World Englishes. As for the language-internal variable analysis, future work should focus on multifactorial approaches that include the full set of variables that we discussed here, since in these other types of analysis, things might look rather different. Special attention needs to be paid to those variables related to the Complexity Principle, in order to confirm whether an increase in cognitive complexity indeed favours the use of the finite CCs rather than the traditional non-finite variant.

Works Cited

- ANNAMALI, Elay and Sanford B. Steever. 1998. "Modern Tamil." In Steever 1998, 75-99.
- ARNOLD, Jennifer, Renee Blake, Brad Davidson, Scott Schwenter and Julie Solomonm, eds. 1996. *Sociolinguistic Variation: Data, Theory and Analysis. Selected Papers from NWAV 2 at Stanford*. Stanford: Center for the Study of Language and Information.
- BIEWER, Carolin. 2015. *South Pacific Englishes: A Sociolinguistic and Morphosyntactic Profile of Fiji English, Samoan English and Cook Islands English*. Amsterdam and Philadelphia: John Benjamins. <https://doi.org/10.1075/veaw.g52>
- Cambridge Dictionary Online. S.v. "remember." <https://dictionary.cambridge.org/es/diccionario/ingles/remember> [Accessed online on June 27, 2022].
- CARDONA, George and Dhanesh Jain, eds. 2007. *The Indo-Aryan Languages*. New York: Routledge.
- CHANDRALAL, Dileep. 2010. *Sinhala*. Amsterdam and Philadelphia: John Benjamins. <https://doi.org/10.1075/loall.15>
- CUYCKENS, Hubert, Frauke D'hoedt and Benedikt Szmrecsanyi. 2014. "Variability in Verb Complementation in Late Modern English: Finite vs. Non-finite Patterns." In Hundt 2014, 182-203. <https://doi.org/10.1017/CBO9781139507226.014>
- Dalton-Puffer, Christiane, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendler, eds. 2006. *Syntax, Style and Grammatical Norms: English from 1500-2000*. Bern: Peter Lang. <https://doi.org/10.3726/978-3-0351-0272-7>
- DASGUPTA, Probal. 2007. "Bangla." In Cardona and Jain 2007, 351-390.
- DAVIES, Mark. 2013. *Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries (GloWbE)*. <http://corpus.byu.edu/glowbe>. [Accessed online on June 27, 2022].
- DE SMET, Hendrik. 2013. *Spreading Patterns: Diffusional Change in the English System of Complementation*. Oxford: Oxford UP. <https://doi.org/10.1093/acprof:oso/9780199812752.001.0001>

- EBERHARD, David M., Gary F. Simons and Charles D. Fennig, eds. 2019. *Ethnologue: Languages of the World*. <http://www.ethnologue.com>. [Accessed online on June 27, 2022].
- FANEGO, Teresa. 1996. "On the Historical Development of English Retrospective Verbs." *Neuphilologische Mitteilungen* 97: 71–79.
- FERNÁNDEZ-SAN MIGUEL, María and Claus-Peter Neuman, eds. 2018. *Taking Stock to Look Ahead: Celebrating Forty Years of English Studies in Spain*. Zaragoza: Prensa Universitaria de la Universidad de Zaragoza. <https://doi.org/10.26754/uz.978-84-16723-51-5>
- FIELD, Andy P. 2009. *Discovering Statistics Using SPSS: (and Sex and Drugs and Rock 'n' Roll)*. 3rd ed. Los Angeles: SAGE Publications.
- FRAMENET. "The Berkeley FrameNet Project." <https://framenet.icsi.berkeley.edu/fndrupal/> [Accessed online on June 27, 2022].
- GAIR, James W. 2007. "Sinhala." In Cardona and Jain 2007, 766–817.
- GARCÍA-CASTRO, Laura. 2018. "Variation in Verbal Complementation Patterns in Nativised Varieties of English: The Case of REMEMBER in Indian English." In Fernández-San Miguel and Neuman 2018, 207-213.
- GARCÍA-CASTRO, Laura. 2020. "Variability in Verbal Complementation in Postcolonial Englishes: Finite versus Non-finite Clauses Governed by REMEMBER." *World Englishes* 39: 411-426. <https://doi.org/10.1111/weng.12481>.
- HASSELGREN, Angela. 1994. "Lexical Teddy Bears and Advanced Learners: A Study into the Ways Norwegian Students Cope with English Vocabulary." *International Journal of Applied Linguistics* 4: 237–58. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- HUDDLESTON, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge UP. <https://doi.org/10.1017/9781316423530>
- HUNDT, Marianne. 2014. *Late Modern English Syntax*. Cambridge: Cambridge UP. <https://doi.org/10.1017/CBO9781139507226>
- KACHRU, Braj B. 1982. *The other tongue: English across cultures*. Urbana: University of Illinois Press.
- KACHRU, Yamuna. 2006. *Hindi*. Amsterdam: John Benjamins. <https://doi.org/10.1075/loall.12>
- KRISHNAMURTI, Bhadriraju. 1998. "Telugu." In Steever 1998, 202-240. <https://doi.org/10.1006/viro.1997.8908>
- KORTMANN, Bernd and Benedikt Szmrecsanyi, eds. 2012. *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin: Mouton De Gruyter. <https://doi.org/10.1515/9783110229226>

- KYTÖ, Merja, Mats Rydén and Erik Smitterberg, eds. 2006. *Nineteenth-century English: Stability and Change*. Cambridge: Cambridge UP. <https://doi.org/10.1017/CBO9780511486944>
- LOUREIRO-PORTO, Lucía. 2017. "ICE vs GloWbE: Big Data and Corpus Compilation." *World Englishes* 36: 448-470. <https://doi.org/10.1111/weng.12281>
- MAIR, Christian. 2006. "Nonfinite Complement Clauses in the Nineteenth Century: The Case of *remember*." In Kytö, Rydén and Smitterberg 2006, 215-228. <https://doi.org/10.1017/CBO9780511486944.009>
- Oxford Dictionaries Online*. S.v. "remember."
<https://en.oxforddictionaries.com/definition/remember> [Accessed online June 27, 2018]
- ROHDENBURG, Günter. 1996. "Cognitive Complexity and Increased Grammatical Explicitness in English." *Cognitive Linguistics* 7: 149-182. <https://doi.org/10.1515/cogl.1996.7.2.149>
- ROHDENBURG, Günter. 2006. "The Role of Functional Constraints in the Evolution of the English Complement System." In Dalton-Puffer, Kastovsky, Ritt and Schendler 2006, 143-166.
- SCHNEIDER, Edgar W. 2013. "English as a Contact Language: The "New Englishes"." In Schreier and Hundt 2013, 131-148. <https://doi.org/10.1017/CBO9780511740060.008>
- SCHNEIDER, Edgar W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge UP. <https://doi.org/10.1017/CBO9780511618901>
- SCHREIER, Daniel and Marianne Hundt, eds. 2013. *English as a Contact Language*. Cambridge: Cambridge UP. <https://doi.org/10.1017/CBO9780511740060>
- STEEVER, Sanford B., ed. 1998. *The Dravidian Languages*. London and New York: Routledge.
- STEGER, Maria and Edgar W. Schneider. 2012. "Complexity as a Function of Iconicity: The Case of Complement Constructions in New Englishes." In Kortmann and Szmrecsanyi 2012, 156-191. <https://doi.org/10.1515/9783110229226.156>
- THOMASON, Sarah G. 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh UP. <https://doi.org/10.1016/B0-08-043076-7/03032-1>
- THOMPSON, Hanne-Ruth. 2012. *Bengali*. Amsterdam: John Benjamins.
- VISSER, Federicus Theodorus. 1963-1973. *An Historical Syntax of the English Language*. 4 parts. Leiden: E.J. Brill.
- WALD, Benji. 1996. "Substratal Effects on the Evolution of Modals in East Los Angeles English." In Arnold, Blake, Davidson, Schwenter and Solomon 1996, 513-530.
- WILLIAMS, Jessica. 1987. "Non-native Varieties of English: A Special Case of Language Acquisition." *English World-Wide* 8: 161-199. <https://doi.org/10.1075/eww.8.2.02wil>