

Exploring variation in English as a lingua franca: Multivariate analysis of modal verbs of obligation and necessity in the VOICE corpus

Research Article

Chunyuan Nie

University of Eastern Finland

Received 22 December 2023; accepted 2 February 2024

Abstract: The modal verbs of necessity and obligation, a testing ground of grammatical change, have been shown to exhibit change and variation in world Englishes. Previous studies have primarily concentrated on English as a native language (ENL) and English as a second language (ESL) varieties. The present study extends this line of research and explores variation in modal verbs of necessity and obligation in English use as a Lingua Franca (ELF). Descriptive statistics indicate that ELF resembles American English and also shares similarities with ESL varieties. In addition, ELF further exhibits divergence from both ENL and ESL varieties that arises in multilingual interactions. The multivariate analysis of this study employs mixed-effects logistic regression on the use of *must* and *have to*. Integrating social and linguistic factors, this analysis exploits metadata gathered from the VOICE corpus, which has thus far been underused. The results of the inferential statistics indicate that the same sociolinguistic factors that influence the variation in ENL and ESL varieties also shape ELF grammar. These findings not only bring ELF closer to other English varieties but also demonstrate the advantage of studying ELF from a variationist sociolinguistic perspective.

Keywords: *grammatical change • ELF • sociolinguistics • modal verbs • mixed-effects • logistic regression*

© Sciendo

1 Introduction

Comprehensive quantitative analysis of grammatical variability in English as a lingua franca (ELF) is still at initial stages, although such approaches are extensive in both English as a native language (ENL) and English as a second language (ESL) varieties. The present study fills in the research gap by investigating the use of modal verbs of necessity and obligation in ELF with descriptive and inferential statistics. The core modal *must* coexists with the semi-modals *have to*, *need to*, and *(have) got to* to express obligation and necessity. The consensus from earlier corpus-based studies reveals a broad quantitative trend in which *must* is declining in frequency, while the uses of *have to* and *need to* are increasing in expressions of obligation and necessity (Krug 2000; Collins 2009; Leech 2013). Previous research suggests that such a frequency shift is attested in both ENL and ESL varieties, with variation observed not only between ENL and ESL varieties, but also within each of the two varieties (Collins 2005; Biewer 2011; Loureiro-Porto 2016). To account for the mechanisms underlying grammatical change and variation, prior studies have investigated the influence of a range of social and linguistic factors using inferential statistics. These studies reveal parallel influences of these factors on the choice of these near synonyms in both ENL and ESL varieties (Tagliamonte and D'Arcy 2007; Tagliamonte and Denis 2014; Hansen 2018). The results suggest that (i) the linguistic factors, such as animacy and the grammatical person of the subjects, as well as social factors like the age and gender of the speaker – which influence ENL varieties – also exert influence on ESL varieties; (ii) different English varieties are at different stages in the process of grammatical change. This study extends the scope of the research by including ELF.

* Corresponding author: Chunyuan Nie, E-mail: chunyuan.nie@uef.fi

Open Access. © 2024 Chunyuan Nie, published by Sciendo.

 This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The use of ELF, which transcends geographic borders and involves speakers with diversified linguistic and cultural backgrounds, poses a challenge to linguistic research due to its heterogeneity. This heterogeneity does not, however, change its nature as a natural language. As a contemporary linguistic phenomenon, ELF is “a new phenomenon in degree but not in kind” (Widdowson 2017: 101). Earlier studies have detected some recurring features in ELF grammar, such as zero 3rd person singular marking, non-standard use of *who* and *which*, and the omission of *be* (Seidlhofer 2004; Dewey 2007; Kirkpatrick 2013). With the availability of ELF corpora, quantitative analysis of ELF grammar is emerging, although it remains scarce. Moving away from merely listing grammatical features, Laitinen (2016, 2020) provides empirical evidence that the use of ELF accelerates the diffusion of change in modal verbs and calls for advanced quantitative methods to further investigate grammatical variation. While her study focuses on a different aspect from modal verbs, Deshors (2020) is one of the first to employ multivariate analysis on ELF grammar. Her results show that the factors that influence the choice between the two syntactic constructions (verb-particle-object vs. verb-object-particle) in ENL and ESL varieties also impact the choice in ELF. Building on previous empirical studies of necessity modals and ELF grammar, the present study offers a quantitative analysis of modals of obligation and necessity in ELF. It first presents the frequencies and distributions of the modal variants, and then inferential statistics quantify the influence of both social and linguistic factors on the choice between the necessity modals. The empirical part addresses two research questions:

- (i) To what extent do ELF speakers follow the grammatical change in necessity modals observed in ENL and ESL varieties?
- (ii) To what extent do the social and linguistic factors that influence the choice of necessity modals in ENL and ESL varieties also contribute to ELF grammar?

The article is structured as follows. Section 2 offers a brief review of previous research on modal verbs in ENL and ESL varieties, followed by related studies in ELF. Section 3 introduces the data and methodology employed. Section 4 presents the results, and Section 5 discusses the findings and implications of this study.

2 Preliminaries

2.1 Variability in necessity modal verbs and world Englishes

The use of modal verbs has been a recurrent topic in English linguistics for their variability in form and in meaning (Ilse, Cappelle and Hilpert 2023: 4). To start with, the modal verb *must* coexists with the semi-modal verbs *have to*, *need to*, *(have) got to* to encode obligation and necessity (root modality), as in examples (1)–(4):¹

- (1) and we **must** calculate that we need three months of lead time to get the goods (VOICE, PBmtg463:538)
- (2) so i'll i'll **have to** remove my microphone afterwards (VOICE, EDcon250:304)
- (3) you guys you guys **need to** go home and sleep and rest yourself (.) (VOICE, EDcon496:261)
- (4) well you'll probably <@> (gonna) **got to** learn some german </@> first (VOICE, LEcon562:1830)

Historically, *must* is the oldest form and came to encode root modality approximately 1,000 years ago (Traugott and Dasher 2002). With *have to*, *have got to* and *need to* emerging successively, *must* has not been discarded but coexists in expressing necessity and obligation. The variant forms are thus layered (Hopper 1991). The coexistence and interchangeability of necessity modals in certain contexts (Flach, Cappelle and Hilpert 2023: 150) exemplify variability in their form. In addition, necessity modals have undergone semantic expansion to epistemic meanings expressing an inference/possibility, as in (5):

1 The transcription conventions observed by VOICE are available at <https://voice2.acdh.oeaw.ac.at/index.xml> (last accessed 31 October, 2023).

- (5) the people **must** be think <6> what these students </6> <7> talking about </7>
 <1> <@> you know? </@> @@ </1> (VOICE, EDcon496:241)

The oldest variant, *must*, developed epistemic senses in the Middle English period (Traugott and Dasher 2002). Although the overall frequencies of epistemic uses are infrequent, it is *must* that is taking the leading role in expressing possibility, but such uses are rarely detected in *have to*, *need to*, and *(have) got to* (cf. Collins 2009; Hansen 2018: 11; Loureiro-Porto 2019). The frequencies of necessity modals and their epistemic meanings are the two primary indicators of variability in these empirical studies.

As mentioned in Section 1, a general trend in the variability of necessity modals is the decrease of *must* and the rise of semi-modals in use. Krug (2000) proposes that the emergence of semi-modals is the result of grammaticalization in which lexical words gain grammatical properties. The decline of *must* is perceived as a sign of democratization, given its association with will imposing from authority (Leech 2003: 237). Moreover, two broadly observed patterns of frequency shift – Americanization towards American English (AmE) norms, and colloquialization towards spoken conventions — are both reflected in the increase of semi-modals and decrease of *must* (Leech 2003: 236). That is, compared with other Englishes, AmE leads the frequency shift with a lower *must* frequency (in its root meaning) and higher semi-modal frequencies. Additionally, the change is more advanced in spoken language than in written modes.

While Krug (2000) presents the increase in semi-modals in British English (BrE) and AmE, Leech (2003, 2011), Smith (2003), and Leech and Smith (2009) provide empirical evidence for the decline of *must* in the last century in BrE and AmE. Collins (2005) extends the study of modal verbs to New Zealand English (NZE) and Australian English (AusE). His observations show that, with the lowest frequency of *must* and the highest frequencies of *have to* and *need to*, AmE is leading the way in the grammatical change in ENL varieties (Collins 2005: 253). In addition, although the frequencies of *must* vary across the four native Englishes that are explored, the root *must* is consistently more favored in written than spoken language (Collins 2005: 254). Collins (2009) provides evidence for the decline of *must* and increase in the semi-modals in ESL varieties, although this trend is comparatively weaker. He argues that stylistic variation exists within both ENL and ESL varieties (2009: 285–287). In the ESL varieties, for instance, Indian English (IndE) is conservative in terms of the overall frequencies the semi-modals but turns out to be the most advanced in terms of colloquialization (cf. Collins and Yao 2012). Biewer (2011) observes that ESL varieties show greater variability in the use of (semi-)modals than ENL varieties do. For instance, the frequencies of *must* range from 5.4 to 6.1 per 10,000 words in ENL varieties, while they range from 6.1 to 9.6 in ESL varieties (Biewer 2011: 22). Applying Second Language Acquisition (SLA) theory, she attributes the variability in ESL varieties to SLA constraints, cultural norms, and transfer from learners' native languages (Biewer 2011: 27–28). Loureiro-Porto (2016) applies Schneider's Dynamic Model (2007) to compare necessity modals in BrE, IndE, and Hong Kong English (HKE). She argues that ESL varieties, at different stages in the formation of new varieties, are consequently at different stages of grammatical change in the use of necessity modals (Loureiro-Porto 2016: 168). Additionally, in a study that also includes IndE and HKE, Hansen (2018: 239–253) shows that ESL varieties present higher frequencies of both *must* and *have to* than ENL varieties, which she explains with cultural factors: a different construal of self as independent or interdependent in different cultures may have an impact on how the modal verb *must* is interpreted.

In investigating what drives speakers to choose one (semi-)modal over another when expressing necessity or obligation, several previous multivariate analyses demonstrate the influence of both linguistic and social factors (Jankowski 2004; Tagliamonte 2004; Tagliamonte and D'Arcy 2007; Tagliamonte and Denis 2014; Hansen 2018; Flach, Cappelle and Hilpert, 2023). These studies include a range of linguistic factors as potential factors that condition observed variation, including the animacy, grammatical person, and reference of the subjects, as well as the type of verbs – all associated with the strength of the obligation as originally hypothesized by Coates (1983). Specifically, according to the cline theory of Coates (1983: 34), second grammatical person/animate/non-generic subjects and dynamic verbs signal stronger expression of obligation. At its simplest, the more specific the necessity/obligation is, the stronger it is. In interpreting the results of multivariate analyses, for instance, when the oldest

variant, *must*, is preferred with factor levels that are associated with weak obligation (third grammatical person, stative verb, etc.), it is indicative of an advanced stage in the process of grammatical change.

In a recent study that demonstrates the advanced stage of AmE, Flach, Cappelle and Hilpert (2023) show that *must* favors 3rd grammatical person and generic subjects in AmE, which suggests that *must* occurs more often in contexts where the speaker has less control (Flach, Cappelle and Hilpert 2023: 160). Additionally, both Hansen (2018) and Flach, Cappelle and Hilpert (2023) incorporate an external factor, text type/genre, as an indicator of formality in their multivariate analysis. The results indicate that it significantly influences the speakers' choice between *must* and *have to* in both ENL and ESL varieties, with *must* favored in more formal contexts.

In addition to the linguistic factors, some studies have adopted a variationist sociolinguistic approach and have included speakers' age and gender as factors. The assumption has been that young speakers and females tend to act as leaders of language change (Labov 1994; Labov 2001). When male/older speakers tend to use the older and more conservative variant *must* more frequently (or female/younger speakers use *have to* more often), the conclusion has been that these quantitative patterns imply ongoing change. To illustrate, Tagliamonte (2004) shows that the use of root *must* has decreased markedly in the apparent time construct (from the oldest to the youngest age group) in York English. Additionally, the use of *have to* is not influenced by age but shows a preference for stative verbs, which, she suggests, signifies the sustained original lexical meaning of *have* (2004: 50). Using the apparent time method, Tagliamonte and D'Arcy (2007) show that the increase in the use of *have to* from older towards younger speakers is obvious in Canadian English, and that females prefer *have to* while males disfavor it (Tagliamonte and D'Arcy 2007: 78). Consequently, while York English is considered conservative, Canadian English is "on the forefront of change" (Tagliamonte and D'Arcy 2007: 83). Including a range of ENL and ESL varieties, Hansen (2018: Chapter 9) assumes that the English varieties are in the same direction in the process of modal system change, but at different speeds. For instance, age, which suggests ongoing change, influences the choice between *must* and *have to* in HKE and IndE, whereas it does not impact BrE (Hansen 2018: 291–295).

In summary, the variation in the frequency shifts of necessity modals both between and within ENL and ESL varieties has been studied from sociolinguistic perspectives. In terms of the influence of social and linguistic factors, the results indicate that various English varieties are at different stages of change.

2.2 Grammatical variation in ELF

The quantitative patterns across ENL and ESL varieties provide evidence of the decline in the use of *must* and the increase in the corresponding semi-modals. Corpus-based observations suggest that diverse social and linguistic backgrounds result in more heterogeneity within ESL than in ENL varieties. Similar to ESL speakers, ELF speakers are mostly non-native (Mauranen 2018a: 107). However, unlike ESL speakers, ELF speakers typically do not share a first language or cultural background (Mauranen 2012: 256). In addition, ELF and ESL varieties are shaped in different socio-political contexts: ESL varieties are the result of colonization, while ELF intertwines with globalization. Currently, ELF is not undergoing indigenization in any country or region, and its communities are not going through the stages of identity reconstruction, such as the ones described in Schneider (2007). Instead, ELF is situation-based, and its communities range from transient occasions of service encounters to permanent relationships, such as transnational marriages (Mauranen 2018b). Despite its heterogeneity, ELF shows some consistent features, as mentioned in Section 1, which are also present in ESL varieties. Jenkins (2009: 201) proposes that both ELF and ESL varieties are products of language contact, involving processes like regularization, leading to parallel surface features. Schneider (2012) presents empirical evidence demonstrating the overlapping features in ELF and ESL varieties. He suggests that these similarities stem from parallel cognitive processes or principles, as both ELF and ESL varieties are language systems influenced by linguistic contact (Schneider 2012: 63). In brief, as much uncertainty and heterogeneity as ELF embraces, it does not necessarily mean that grammatical change is impossible to detect in ELF. As Seidlhofer (2011: 70) argues, being a natural language, ELF can be compared to both ENL and ESL varieties, and the variation resulting from its unprecedented spread and diverse contexts should be of interest to sociolinguistics.

Given that ELF takes place in diverse and heterogeneous settings, it is not surprising that researchers have started to explore variation that emerges in it. Using descriptive statistics, Laitinen (2016, 2020) investigates modal verbs in both written and spoken ELF corpora. He compares the frequencies of a set of (semi-)modals across ELF, ENL, and ESL varieties, and the empirical evidence suggests that ELF follows AmE in the process of the ongoing change, and that ELF users are more advanced than those of ESL varieties in the use of both full and semi-modals (Laitinen 2016: 193). To explain the finding, he turns to the social characteristics of ELF users. By definition, ELF users are multilingual and tend to have a greater number of weak social connections compared to speakers of other English varieties (Laitinen and Lundberg 2020). Given Milroy and Milroy's (1985) observation that change typically spreads through weak tie environments, it is likely that ELF users contribute to accelerating ongoing changes. The association between ELF users' social characteristics and linguistic change necessitates further exploration using advanced statistical methods, as indicated by Laitinen (2020) and Deshors (2020). This association also emphasizes the value of a sociolinguistic perspective in studying ELF.

By addressing the two research questions in Section 1, the present study contributes to ELF research and the study of ENL and ESL varieties, as well as demonstrates the advantages of studying ELF from a variationist sociolinguistic perspective.

3 Data and methodology

3.1 Data and the variants

The empirical evidence comes from The *Vienna-Oxford International Corpus of English* (VOICE.2013), a spoken ELF corpus of approximately one million words of naturally occurring face-to-face conversations. VOICE contains data compiled between 2001 to 2007, obtained from leisure, educational, and professional domains. The corpus covers a wide array of speech event types, such as meetings, interviews, conversations, service encounters, discussions, panels, question-answer sessions, and press conferences.

For the purpose of this study, the dependent variables are extracted from VOICE texts with AntConc (Anthony 2020) and are then manually coded. To compare the frequencies in ELF with ENL and ESL, this study first presents a bird's-eye view of the four variants *must*, *have to*, *need to*, and *(have) got to* with descriptive statistics. Secondly, inferential statistics focus on the alternation between the uses of root *must* and *have to*. While *must* is the older variant, *have to* is the currently dominant semi-modal expressing obligation and necessity in both ENL and ESL varieties. Semantically, *must* and *have to* are regularly paired to demonstrate a contrast: subjectivity vs. objectivity (Smith 2003: 243). While *must* generally involves will imposed directly by the speaker (authority), *have to* mostly relates to expressing objective necessity, such as laws and regulations (Depraetere and Verhulst 2008). Methodologically, this study extends the exploration of necessity modals to ELF, focusing on the choice of these variants. In previous multivariate analyses, *must* and *have to* are consistently included, with *(have) got to* being included on a less frequent basis (Jankowski 2004; Tagliamonte 2004; Tagliamonte and D'Arcy 2007; Tagliamonte and Denis 2014; Hansen 2018; Flach, Cappelle and Hilpert 2023). However, due to the low frequency of *(have) got to* in VOICE, this study includes only *must* and *have to* in the multivariate analysis, aligning with the studies conducted by Hansen (2018) and Flach, Cappelle and Hilpert (2023). To ensure comparability with earlier studies (e.g. Tagliamonte and D'Arcy 2007; Hansen 2018), the multivariate analysis is limited to contexts of present tense affirmative sentences where the variants are interchangeable, as illustrated in examples (6)–(7):

- (6) that this is not (.) an (adequate) procedure and (.) i **must** then explain as i best can that it's workable under the circumstances = (VOICE, POMtg542:308)
- (7) you **have to** be aware of the whole system and differences hh (VOICE, POMtg314:25)

While *must* does not have a past tense, it is different from *have to* in expressing negation and interrogation. Occurrences with negation, interrogation and past tense, such as in (8)–(9), have been excluded:

- (8) the tests you **don't have to** answer those questions on <8> the </8> exam (VOICE, EDcon521:932)
 (9) we **had to** write an (.) intelligence test? or something similar (VOICE, EDcon250:904)

Additionally, occurrences with repetition and unfinished contexts have also been excluded:

- (10) you **must** hh er (.) you **must** <LNger> ke- e:r kennen {know} </LNger> (.) <soft> you **must** <LNger> kennen {know} </LNger></soft> (.) <9> you: should? </9> (VOICE, EDsed251:373)

As in (10), there are three occurrences of *must*, but only one is included. Unfinished utterances are only found in the case of unidentified speakers and have been automatically removed in the multivariate analysis.

3.2 Independent variables

The independent variables in the multivariate analysis include both linguistic and social factors, shown in Table 1.

Table 1. Social and linguistic factors.

Factors	Levels
Grammatical person	1 st , 2 nd , 3 rd
Animacy	animate, inanimate
Reference	non-generic, generic
Verb	dynamic, stative
Age	17–24, 25–34, 35–49, 50+
Gender	male, female
Power relation	symmetric, asymmetric

First, the study includes the four linguistic factors that have been shown to influence variation in ENL and ESL: grammatical person, animacy, reference of the subjects, and the verb type. The distinctions between the variable levels are illustrated in examples (11)–(13):

- (11) we **have to have** a solution (.) (1st, animate, non-generic, stative verb; VOICE, EDcon521:330)
 (12) **the killing of innocent civilians must stop** (.) (3rd, inanimate, non-generic, dynamic verb; VOICE, POprc522:5)
 (13) because in my country for example (.) y- you you can't be (.) hh to be a doctor **you must be** a doctor (2nd, animate, generic, stative verb; VOICE, POWgd14:1358)

In example (13), the second person, animate subject, usually considered as non-generic in a conversation, is instead regarded as generic rather than non-generic. This is because, in cross-cultural interactions in VOICE, it is typical for a speaker to introduce a widespread situation or a general rule in their country. The distinctions between the variable levels are generally clear. When the subjects are organizations, corporations, and institutions, they are coded as inanimate unless they clearly refer to a group of people. When the subject or the verb is unclear, it will be left unknown unless it can be identified in the context.

Second, the rich metadata in VOICE allows us to investigate the possible influence of external social factors in the alternation between (semi-)modals. In terms of social factors, age, gender,² and power

2 This study directly adopts the gender distinction presented in metadata, separating speakers into males and females.

relations are included. The influence of age and gender on linguistic change and variation has been extensively discussed in the context of world Englishes, where females and young speakers are often seen as drivers of grammatical change (Fuchs and Gut 2015; Tagliamonte 2016; Suárez-Gómez and Seoane 2023, etc.). Specifically, Tagliamonte (2004), Tagliamonte and D'Arcy (2007) and Hansen (2018) attest to the influence of age and gender in the modal verbs of ENL and ESL varieties. ELF speakers are exposed to greater variability in multilingual contexts. This makes it essential to examine the social factors that could unveil the complex relationship between grammatical change and contexts of use (Seidlhofer 2011: 70). In addition to age and gender, the study includes power relations as a potential variable. It can be directly extracted from the metadata. Power relations describe the authority hierarchy between speakers, which is related to the use of *must*. If, for instance, the interaction is between peers, it is considered symmetric, whereas the interaction between professors and students is asymmetric.

3.3 Methodology

The empirical evidence includes descriptive statistics and inferential statistics. The first part reports the overall distributions of the (semi-)modals in VOICE and compares the frequencies with those observed in ENL and ESL. The data for the comparisons in Section 4 have been taken from Hansen (2018: 366). Her findings derive from the spoken part of the *International Corpus of English* (ICE) and the *Santa Barbara Corpus of Spoken American English* (SBC), neither of which shares exactly the same design as that of VOICE.³ The comparison here is primarily intended to provide a general picture of ELF among English varieties. The multivariate analysis scrutinizes the alternation between root *must* and *have to* with mixed-effects logistic regression. To quantify simultaneously the influence of social and linguistic factors on the speakers' choice between *must* and *have to*, binary logistic regression has been the common choice (Levshina 2015). However, VOICE data violates one prerequisite of logistic regression in that the observations are not independent. It is frequent in VOICE that one speaker contributes to several or even dozens of tokens, as in Figure 1.

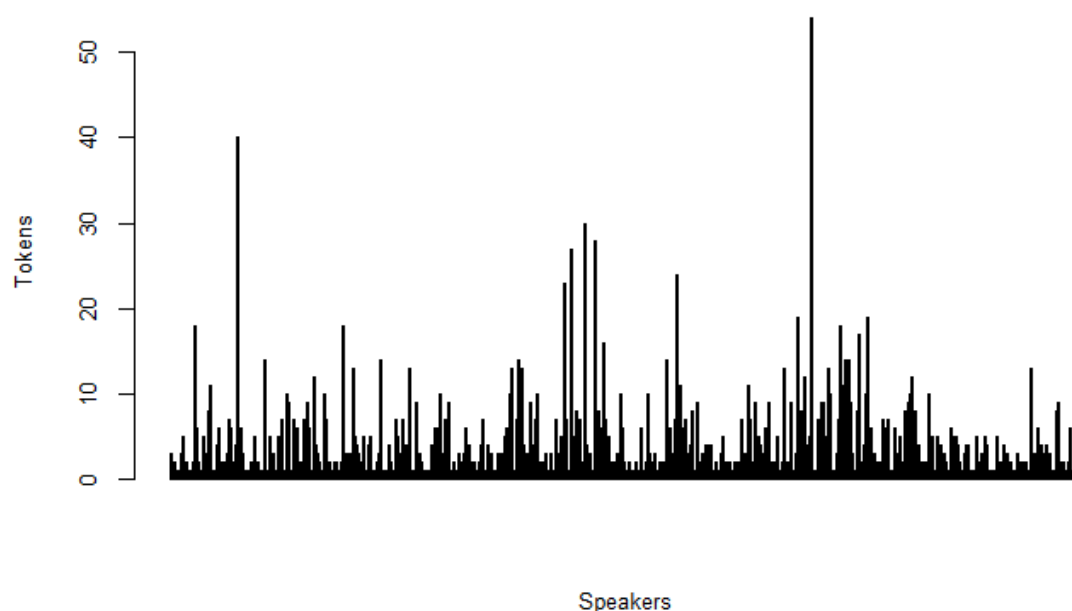


Figure 1. Frequency of (semi-)modals by speaker in VOICE.

3 For instance, while casual conversation makes up 10 percent in VOICE, it constitutes 30 percent in ICE. While VOICE includes conversations only, the spoken part of ICE includes both conversations and monologs. Monologs in ICE include demonstrations, presentations, and speeches, which also exist in VOICE.

Figure 1 shows that many speakers contribute several tokens of the (semi-)modals. Traditional logistic regression models that include only fixed-effect variables are problematic in this situation because the speakers themselves could be a source of variation (Tagliamonte and Baayen 2012; Gries 2015). The fixed effects are constant and repeatable (Winter 2020: 236). For instance, the linguistic and social factors listed in Table 1 are fixed-effect variables since they could be studied repeatedly using different corpora. Conversely, individual speakers, randomly selected from the population, are a typical random-effect variable when one individual produces multiple observations (Tagliamonte and Baayen 2012: 8). A mixed-effects model that takes random effects into consideration “increases the precision of the regression results” (Gries 2015: 97). Random-effect variables can be not only individual specific but also item specific. Traditionally, external factors that categorize text types are analyzed as fixed effects since they show a hierarchy of formality (see Section 2.1). However, two similar factors, domains and speech event types, are categorized as random effects. Firstly, there is a hierarchy of structure between domain and speech event type (Gries 2015: 99). Secondly, these two factors fall into random-effect variables for they do not “exhaust all possible levels of the factor” (Tagliamonte and Baayen 2012: 143). In addition, domains and speech event types in VOICE do not show a hierarchy of formality. In all, the mixed effects logistic regression model includes all the variables in Table 1 as fixed effects and speaker, domain, and speech event type as random effects.

4 Results

4.1 Distribution of necessity modals in ELF

Comparing the overall frequencies (per one million words) of (semi-)modals in VOICE with those in ENL and ESL varieties in Hansen (2018), Figure 2 provides an overview of ELF among English varieties.

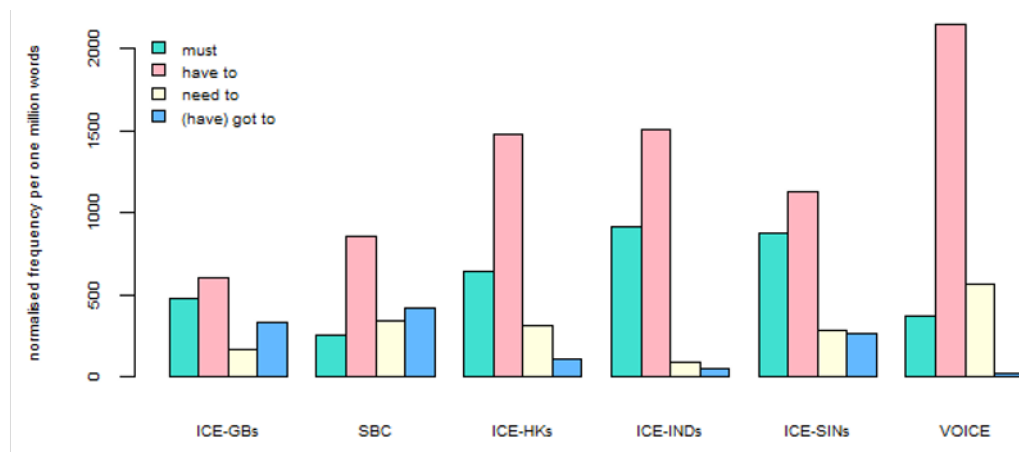


Figure 2. Overall frequencies of *must*, *have to*, *need to*, and *(have) got to* in VOICE, ENL and ESL varieties (data from ENL and ESL varieties based on Hansen 2018).

Figure 2 shows that in the variability of form, specifically in the layering of (semi-)modals, VOICE resembles AmE in the layering of *have to*, *need to*, and *must*. However, with a low frequency of *(have) got to*, ELF falls between the ESL varieties. Overall, the increasing semi-modal *have to* demonstrates the highest frequencies across all varieties, but the declining core modal *must* behaves divergently. When comparing the frequencies of *must* and *have to*, the two contrastively paired variants, VOICE shows the highest ratio (5.8) of *have to*/*must*. The ratio of *have to* to *must* in AmE, which is considered to be the most advanced variety in the process of change, is 3.6. In other words, the contrast between *must* and *have to* is the strongest in VOICE.

As an indicator of advancement in the process of change in this modal category, the proportions of epistemic modality diverge (Figure 3). The overall frequency of *must* is low in VOICE, but when we take a closer look at epistemic *must*, ELF displays a conservatism that resembles the ESL varieties.

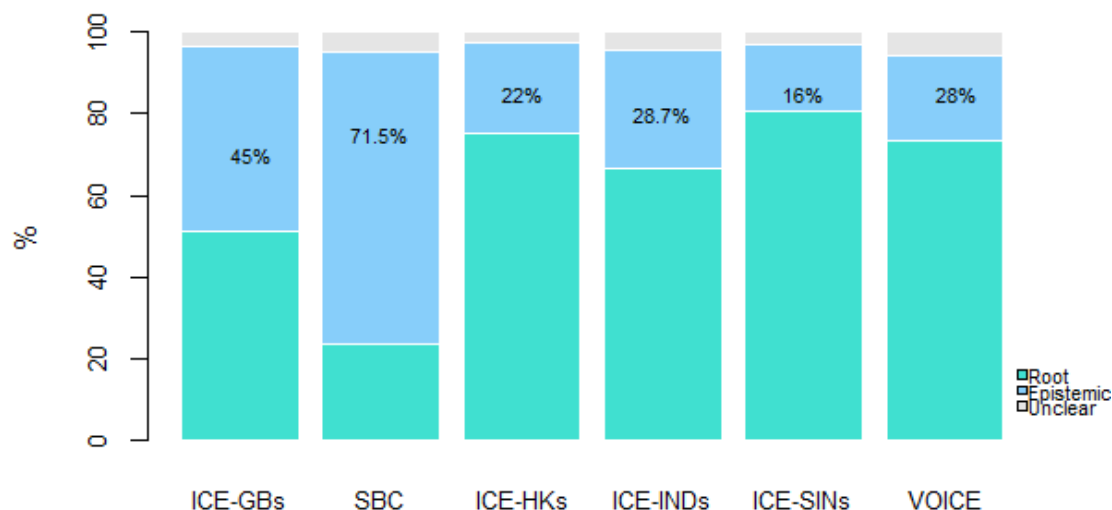


Figure 3. Distributions of root and epistemic *must* in VOICE, ENL and ESL varieties (data from ENL and ESL varieties based on Hansen 2018).

The proportion of root *must* is high in VOICE, which differs from ENL and falls between the ESL varieties. The low frequency of both *must* and its epistemic meaning in VOICE are contradictory. To illustrate, in ENL and ESL varieties, a low frequency of *must* is typically accompanied by a low proportion of its root meaning, and vice versa, a high frequency of *must* is typically accompanied by a high proportion of its root meaning. As described in Section 3.2, most interactions in VOICE occur in cross-cultural events where speakers introduce widespread situations or general rules, with most uses of root *must* also being observed in those circumstances:

- (14) in order to: be able to change it (.) **we** er **must** (.) t- e:r we mu- er we **must be aware** (.) of the faults the system has (.) (VOICE, EDsed251:524)
- (15) <7> it should be for the entire </7> program <1> and </1> then **you must copy** it a:nd and send it to to the various offices (VOICE, POWgd317:1247)
- (16) before they be able to approve the budget in the parliament (.) **they must first have** the <pvc> approvemnt (VOICE, POWgd510:391)

According to Depraetere and Verhulst (2008), these obligations could be considered as expressing external obligation stemming from the circumstances, regulations, or conditions. This observation implies that, despite its high proportion of root meaning, *must*, even with a second person subject as in (15), is not commonly associated with strong obligation in VOICE.

At this stage, the corpus-based evidence from VOICE shows that ELF speakers clearly follow the grammatical change attested in ENL and ESL. Nevertheless, ELF also shows divergence that requires further analysis. Table 2 and Table 3 illustrate the distributions of root *must* and *have to* according to the linguistic variable levels under investigation.⁴

4 Each level's percentage of the total variable is shown in brackets.

Table 2. Distribution of necessity modals in their root meaning by linguistic factor levels in VOICE.

	Animacy		Reference		Grammatical person			Type of verb	
	animate	inanimate	non-generic	generic	1st	2nd	3rd	dynamic	stative
<i>must</i> (241)	130 (54 %)	111 (46 %)	160 (66 %)	81 (34 %)	68 (28 %)	38 (16 %)	135 (56 %)	140 (58 %)	101 (42 %)
<i>have to</i> (1769)	1661 (94 %)	108 (6 %)	1356 (77 %)	413 (23 %)	841 (48 %)	655 (37 %)	273 (15 %)	1484 (84 %)	285 (16 %)

Table 2 shows that, with the exception of the grammatical person, the linguistic factor levels that are associated with strong obligation – animate, non-generic, and dynamic – dominate the expression of necessity and obligation in *must* and *have to*. Moving beyond this general observation, a closer examination of the individual factors reveals more details. Firstly, a notable pattern is observed in the use of inanimate subjects with *must*. While inanimate subjects are generally infrequent, they are more commonly associated with *must* than with *have to*. This trend is also observed in both ENL and ESL varieties (Collins 2009; Loureiro-Porto 2016). Secondly, the distribution of grammatical person in VOICE also aligns with the findings in Collins (2009) and Loureiro-Porto (2016), showing that the third-person subjects dominate the expressions of root modality with *must*. Lastly, there is no obvious distinction between *must* and *have to* in the distribution of non-generic or generic subjects, but stative verbs are proportionally more frequent with *must* than with *have to*.

Table 3. Distribution of necessity modals in their root meaning by gender and age in VOICE.

	Gender		Age group			
	male	female	17–24	25–34	35–49	50+
<i>must</i> (241)	140 (58%)	101 (42 %)	79 (33%)	46 (19%)	46 (19%)	70 (29%)
<i>have to</i> (1769)	701 (40%)	1068 (60%)	649 (37%)	332 (19%)	445 (25%)	343 (19%)

The data in Table 3 indicate a preference for *must* among males, while females use *have to* more frequently. In terms of age, the pattern is complex. From the perspective of apparent time construct, *must* and *have to* show similarities: firstly, there is an increase in usage among the youngest age group, followed by a mild decrease in the 25–34 age group. Additionally, there is a slight increase in the usage of *have to* towards the youngest age group, while the 50+ age group shows the second highest proportion in the use of *must*. The distributions presented in Table 2 and Table 3, which suggest the influence of both linguistic and social factors on the choice between *must* and *have to*, will be subjected to advanced statistical analysis to further explore these patterns.

4.2 Multivariate analysis

The multivariate analysis examined 241 tokens of *must* and 1769 of *have to* in the contexts described in Section 3. The mixed-effects logistic regression was conducted with R (v 4.1.2; R Core team 2021). There were originally three random-effect and seven fixed-effect variables to be examined in the mixed-effect model. The low VIF values (< 2) of all the independent variables suggest they are not correlated, and multicollinearity is not a concern here.⁵ In the first place, among the three random-effect variables, domain and speech event type were dropped because the variances they produced were both close

5 See Appendix 1 for the VIF scores of the independent variables.

to zero, which suggests variability is hardly detected within them.⁶ Secondly, the original model, which included all seven fixed-effect variables and speaker as the random effect, was compared to a model containing only the three significant fixed-effect variables, also with speaker as the random effect. Both the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are goodness-of-fit measures to compare if the models have too many variables. Generally, lower AIC and BIC values indicate a better model fit (Levshina 2015). The model with only the three significant fixed-effect variables and speaker as the random effect yielded slightly lower AIC (926.5) and BIC (954.5) values compared to the original model's AIC (933.1) and BIC (994.8). Furthermore, the results of the ANOVA test showed no statistically significant difference between the model with all fixed-effect variables and the model with only the significant variables. In conclusion, the results presented in Table 4 retain all seven fixed-effect variables, along with the inclusion of speaker as the random effect.

Table 4. Results of the mixed-effects regression model.

<i>Predictors</i>	VOICE		
	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	0.00	0.00 – 0.01	<0.001
Person [1st]	1.79	0.88 – 3.63	0.108
Person [3rd]	3.33	1.31 – 8.47	0.011
Verb [stative]	3.63	2.00 – 6.59	<0.001
Animacy [inanimate]	27.45	9.22 – 81.74	<0.001
Reference [G]	1.63	0.82 – 3.22	0.162
Gender [M]	3.94	1.37 – 11.36	0.011
Age25-34	0.48	0.10 – 2.23	0.348
Age35-49	0.64	0.16 – 2.56	0.531
Age [50+]	2.10	0.42 – 10.44	0.364
Power [symmetrical]	0.73	0.17 – 3.14	0.675

The odds ratios in Table 4 compare the odds of two variants: in this case, *must* and *have to*. The baseline category of the two variants in this model is *have to*. That is, an odds ratio greater than 1 suggests that the occurrence of *must* is more likely than *have to* (Levshina 2015). For instance, the odds ratio for the factor Gender is 3.94, which means the odds for males to choose *must* are about 3.94 times as high as those for females. Confidence intervals (CIs) quantify how robust the odds ratios are and should not include zero. The significant variables are in bold when $p < 0.05$.

Of the linguistic variables, animacy and grammatical person of subject, and the type of verb are significant. *Must* prefers inanimate subjects and stative verbs, which are associated with weak obligation. Overall, inanimate subjects are rare with *have to*, while they display 46 percent of occurrences with *must*; items (17)–(19) provide some examples:

- (17) all country **must** er e:r er: must must choose (.) (VOICE, EDsed251:77)
- (18) the criteria are (**there**) **must** be <4> some </4> some structure in our work (.) (VOICE, POMtg315:1039)
- (19) **municipalities** will **have to** fund themselves there is a number of large projects (.) (VOICE, PBpan28:106)

6 This illustrates the situation-based feature of ELF and does not necessarily imply that the register variation is not detectable.

As illustrated previously, root *must* in VOICE is largely associated with external obligation even in the case of a second person subject. Therefore, its preference for inanimate subjects is unsurprising. There are a few occurrences, as in (18), where its subject is grammatical (existential *there*), and the logical subject has been moved further away from the modal verb. The preference for inanimate subjects suggests that *must* is more advanced than *have to* in VOICE in the process of grammaticalization. Loureiro-Porto (2016) reports similar findings in BrE, HKE, and IndE with descriptive statistics: although inanimate subjects are infrequent overall, they show a preference for *must* over other (semi-)modals of necessity across all three varieties (BrE is more advanced in this regard). Comparing with 2nd person (the reference level), *must* prefers 3rd person subjects. The 3rd person subjects, which are more likely to associate with inanimate subjects (as seen in examples (17)–(19)), are considered to express weak obligation. The preference for 3rd person with *must* has also been observed in AmE (Flach, Cappelle and Hilpert 2023). Stative verbs are also generally infrequent (more frequent than inanimate subjects) in VOICE. The following are two examples illustrating *must* and *have to* with stative verbs:

- (20) yes and <6> we **must be** </6> **optimistic** not not ne<7>gative </7> (VOICE, EDwsd15: 411)
 (21) we haven't <8> truly discussed that because in the conclusions </8> we **have to have a solution** (.) (VOICE, EDcon521:330)

According to Coates (1983: 37), stative verbs in general tend to occur with weak obligation. The preference for stative verbs further supports the observation that, although root meaning still dominates *must*, it mainly expresses weak obligation in VOICE. Although *must* shows a preference for inanimate and 3rd person subjects, as well as stative verbs, the preference is most pronounced with inanimate subjects, as shown in Table 4. All of this evidence suggests that *must* is more advanced than *have to* in the process of change, with the difference in the degree of preference existing among the linguistic factors.

Of the social factors – age, gender, and power relations – only gender turns out to be significant in the data. The power relations between the speakers, which are related to the use of *must*, do not affect the choice between *must* and *have to*. There is little corpus-based research on the influence of power relations on necessity modals, making it challenging to ascertain if this lack of influence is specific to ELF or if it pertains to the use of necessity modals more in general. In terms of the other two factors, age does not affect the alternation between *must* and *have to* and only mild change can be observed in Table 3. The fluctuations between age groups could be the result of noisy data and imbalanced distribution of speakers in VOICE. Moreover, the differentiation between genders observed in Table 3 is also statistically significant in the regression model. Gender is seldom included in the study of English modal verbs. Nevertheless, Tagliamonte and D'Arcy (2007) present a similar finding in spoken Canadian English: males show a preference for *must*, but only statistically significant in two age groups (17–29 and 30–59). Hansen (2018) documents that males prefer *must* more than do females in HKE and IndE.

Figure 4 focuses on the four significant variables and shows the probabilities of *must*. It provides further evidence that the grammatical change in ELF is not random. As shown in Figure 4, *must* is more likely to appear when the subjects are 3rd person and inanimate, the verbs stative, and the speakers male. Although males show a general preference for *must* over *have to*, Figure 4 shows some consistency between the two genders: the probabilities of using *must* increase for both genders in contexts associated with weak obligation. In terms of animate subjects that are associated with strong obligation, both genders tend to choose *have to*, regardless of the types of verbs. The differentiation between the two genders is more obvious when the subjects are inanimate. The probabilities increase for both genders to choose *must* when the subjects are inanimate and 3rd person, with an increase more obvious for males.

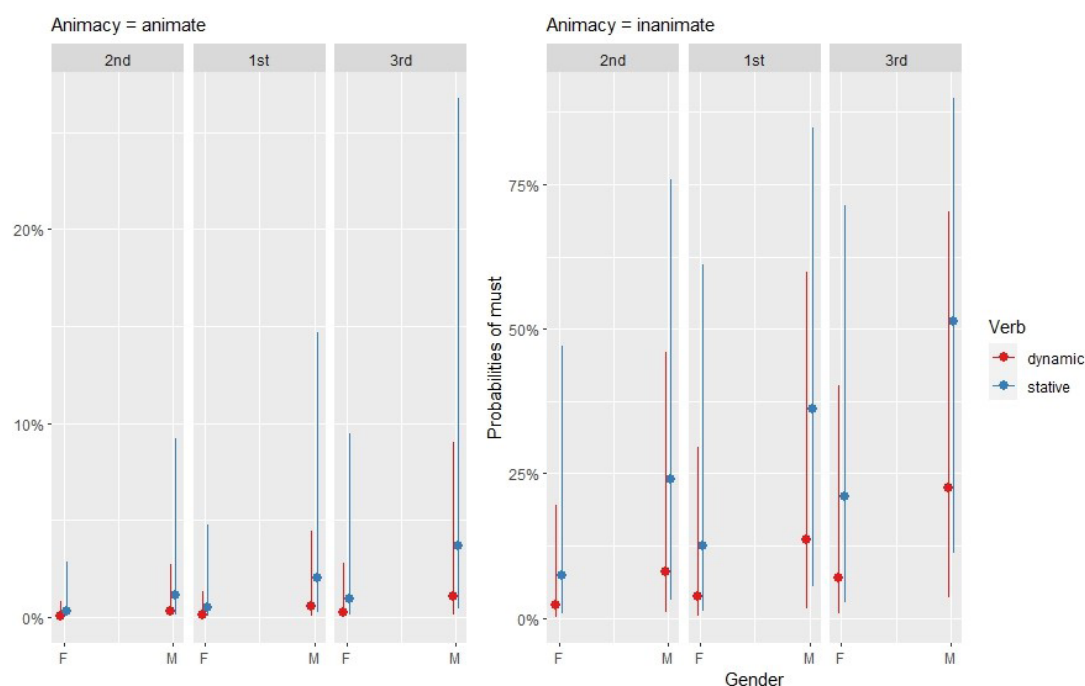


Figure 4. Predicted probabilities of *must*.

The interaction between independent variables has rarely been studied in the field of modal verbs. Nevertheless, this study has tested interactions between the four statistically significant variables. The interaction between the two social factors, age and gender, has also been tested. All the interactions turn out to be insignificant.

5 Discussion and conclusion

This article has explored variation in ELF by investigating modal verbs of obligation and necessity in VOICE. It has utilized both descriptive statistics and multivariate analysis to contribute to the limited existing empirical evidence that linguistic change in ELF is traceable and multifactorial. To answer the two research questions posed in Section 1, the results indicate that (i) ELF speakers also follow the grammatical changes that have been observed and widely verified in ENL and ESL varieties; (ii) the social and linguistic factors that influence the changes in ENL and ESL varieties are also shaping ELF. In terms of the frequency shift, ELF resembles AmE and shows advancement in the grammatical changes with low frequency of *must* and high frequencies of *have to* and *need to*. However, the low proportion of epistemic *must* and the low frequency of *(have) got to* in ELF mirror the trends observed in ESL varieties. These findings suggest that ELF speakers are not only following the leaders of grammatical changes (i.e. AmE) but are also adopting the changes found in other varieties.

In terms of the multivariate analysis that focused on the alternation between *must* and *have to*, the type of verbs, the animacy and grammatical person of subjects, and gender of the speakers are statistically significant in the choice between the two variants. The significant influence of the linguistic variables indicates that although the frequency of root *must* is comparatively high, it is mainly restricted to expressing weak obligation. Additionally, the significant impact of gender suggests that in situation-based ELF interactions women are also more sensitive to linguistic change. The influence of gender could be the result of the fact that while most interactions in VOICE are formal and public, women show a general tendency of sensitivity to social evaluation, especially in formal circumstances (Labov 2001:

291). However, the role of age in grammatical change is less transparent. Hansen (2018: 295) finds that age does not affect the choice between *must* and *have to* in spoken BrE, interpreting this as “the change has reached a saturation point”. This explanation could also apply to the findings of the present study. Another possible explanation for age not influencing the alternation in ELF is communal change. Labov (1994: 84) proposes that communal change takes place when all speakers in a community, regardless of their age, are changing simultaneously and change is not observable at a community level. Nevertheless, the way in which gender or age influence linguistic change is complex and requires further investigation before one can draw conclusions. The statistical significance of the social and linguistic factors used here suggests that the grammatical change is ongoing, and *must* and *have to* are at different stages in the process of change. Similar to Deshors (2020: 10), who has studied particle placement and argues that it “is a multifactorial phenomenon” in ELF, this study indicates that the alternation between modal verbs is also multifactorial. In brief, this study provides evidence that although ELF communities, where English is a dominant lingua franca, are diffuse (Mauranen 2018b: 12), the natural evolutionary processes that influence world Englishes also influence ELF (Laitinen 2020: 440).

ELF not only shows similarities with ENL and ESL varieties but also exhibits its own quantitative patterns. First, the frequency of *have to* is overly high in VOICE compared with other varieties in prior studies. This leads to the strongest contrast, that is, the highest ratio between *have to* and *must*. This finding is similar to an earlier observation by Biewer (2011), who illustrates that ESL varieties show higher frequencies of semi-modals than do ENL varieties and associates this finding with the process of ESL acquisition. Nevertheless, the frequency of *have to* in VOICE is almost twice as high as that in ESL varieties. The highest ratio of *have to/must* could be the result of the social characteristics of ELF speakers in multilingual settings, which accelerate change, as suggested by Laitinen and Lundberg (2020). In addition, it is also possible that the types of interactions included in VOICE, most of which occur in international events and involve the issuing of instructions, influence the frequency of *have to*.

Methodologically, this study demonstrates the potential contribution of the VOICE corpus metadata for sociolinguistic studies of ELF and for future multivariate analyses. Given the fact that such metadata exist, they also need to be used in empirical studies. Moreover, this study adds to the empirical evidence in ELF, demonstrating that grammatical change variation is detectable in ELF and establishes quantitative similarities with ENL and ESL varieties. The findings and discussions here are limited to VOICE, which includes only spoken data and mainly involves European speakers. To unveil the dynamics of grammatical change in ELF, future studies should expand the research by including other ELF corpora, further exploring metadata, and including additional sociolinguistic factors.

Lastly, the discussion of the social and linguistic factors in this study is based on the results of regression models that rely on *p*-values and produce binary answers. Alternatively, future study could analyze the corpus data using Bayesian models. Bayesian models, which do not rely on *p*-values, produce the posterior probabilities of independent variables and permit discussion of subtle differences between data sets (Levshina 2022). Comprehensive quantitative analysis of ELF corpora provides empirical evidence and also contributes a fresh perspective to ELF research.

References

- Anthony, Laurence. 2021. AntConc (Version 3.5.9) [Computer software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Bailey, Guy, Tom Wickle, Jan Tillery and Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3 (3): 241–264.
- Biewer, Carolin. 2011. Modal auxiliaries in second language varieties of English. In J. Mukherjee and M. Hundt (eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap*, 7–33. Amsterdam/Philadelphia: John Benjamins.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Collins, Peter. 2005. The modals and quasi-modals of obligation and necessity in Australian English and other Englishes. *English World-Wide* 26: 249–273.

- Collins, Peter. 2009. Modals and quasi-modals in world Englishes. *World Englishes* 28 (3): 281–292.
- Collins, Peter and Yao Xinyue. 2012. Modals and quasi-modals in new Englishes. In M. Hundt and U. Gut (eds.), *Mapping unity and diversity world-wide. Corpus-based studies of new Englishes*, 35–54. Amsterdam: John Benjamins.
- Depraetere, Ilse and An Verhulst. 2008. Source of modality: A reassessment. *English Language and Linguistics* 12 (1): 1–25.
- Depraetere, Ilse and Susan Reed. 2020. Mood and modality in English. In B. Aarts, A. McMahon and L. Hinrichs (eds.), *The handbook of English linguistics*, 207–227. Hoboken: Wiley-Blackwell.
- Depraetere, Ilse, Bert Cappelle and Martin Hilpert. 2023. Introduction. In S. Fitzmaurice and B. Kortmann (eds.), *Models of modals: From pragmatics and corpus linguistics to machine learning*, 1–13. Berlin/Boston: De Gruyter.
- Deshors, Sandra C. 2020. English as a lingua franca: A random forests approach to particle placement in multi-speaker interactions. *International Journal of Applied Linguistics* 30 (2): 214–231.
- Dewey, Martin. 2007. *English as a lingua franca: An empirical study of innovation in lexis and grammar*. Doctoral dissertation, King's College London.
- Flach, Susanne, Bert Cappelle and Martin Hilpert. 2023. You *must/have* to choose: Experimenting with choices between near-synonymous modals. In S. Fitzmaurice and B. Kortmann (eds.), *Models of modals: From pragmatics and corpus linguistics to machine learning*, 149–176. Berlin/Boston: De Gruyter.
- Gries, Stefan Th. 2015. The most under-used statistical method in corpus-linguistics: Multi-level (and mixed-effects) models. *Corpora* 10 (1): 95–125.
- Hansen, Beke. 2018. *Corpus linguistics and sociolinguistics: A study of variation and change in the modal systems of world Englishes*. Leiden/Boston: Brill.
- Hopper, Paul J. 1991. On some principles of grammaticization. In E. Traugott and B. Heine (eds.), *Approaches to grammaticalization*. Vol. 1: *Focus on theoretical and methodological issues*, 17–35. Amsterdam/Philadelphia: John Benjamins.
- Jankowski, Bridget. 2004. A transatlantic perspective of variation and change in English deontic modality. *Toronto Working Papers in Linguistics* 23: 85–113.
- Jenkins, Jennifer. 2009. English as a lingua franca: Interpretations and attitudes. *World Englishes* 28 (2): 200–207.
- Kirkpatrick, Andy. 2013. The Asian corpus of English: Motivation and aims. *Learner Corpus Studies in Asia and the World* 1: 17–30.
- Krug, Manfred G. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin/New York: De Gruyter.
- Labov, William. 1963. The social motivation of a sound change. *Word* 19 (3): 273–309.
- Labov, William. 1994. *Principles of linguistic change*. Vol. 1: *Internal factors*. Hoboken: Wiley-Blackwell.
- Labov, William. 2001. *Principles of linguistic change* Vol. 2: *Social factors*. Hoboken: Wiley-Blackwell.
- Laitinen, Mikko. 2016. Ongoing changes in English modals: On the developments in ELF. In O. Timofeeva, A. Gardner, A. Honkapohja and S. Chevalier (eds.), *New approaches in English linguistics: Building bridges* (Studies in language companion series 177), 175–196. Amsterdam/Philadelphia: John Benjamins.
- Laitinen, Mikko. 2020. Empirical perspectives on English as a lingua franca (ELF) grammar. *World Englishes* 39 (3): 427–442.
- Laitinen, Mikko and Jonas Lundberg. 2020. ELF, language change and social networks: Evidence from real-time social media data. In A. Mauranen and S. Vetchinnikova (eds.), *Language change: The impact of English as a lingua franca*, 179–204. Cambridge: Cambridge University Press.
- Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, M. Krug and F. Palmer (eds.), *Modality in contemporary English*, 223–240. Berlin/New York: De Gruyter.
- Leech, Geoffrey. 2011. The modals ARE declining: Reply to Neil Millar's "Modal verbs in TIME: Frequency changes 1923–2006", *International Journal of Corpus Linguistics* 14 (2), 191–220 (2009). *International Journal of Corpus Linguistics* 16 (4): 547–564.
- Leech, Geoffrey and Nicholas Smith. 2009. Change and constancy in linguistic change: How grammatical usage in written English evolved in the period 1931–1991. In A. Renouf and A. Kehoe (eds.), *Corpus linguistics: Refinements and reassessments*, 173–200. Leiden/Boston: Brill.

- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.
- Levshina, Natalia. 2022. Comparing Bayesian and frequentist models of language variation: The case of *help* + (to-)infinitive. In O. Schützler and J. Schlüter (eds.), *Data and methods in corpus linguistics – Comparative approaches*, 224–58. Cambridge: Cambridge University Press.
- Loureiro-Porto, Lucía. 2016. (Semi-)modals of necessity in Hong Kong and Indian Englishes. In E. Seoane and C. Suárez-Gómez (eds.), *World Englishes: New theoretical and methodological considerations*, 143–172. Amsterdam/Philadelphia: John Benjamins.
- Loureiro-Porto, Lucía. 2019. Grammaticalization of semi-modals of necessity in Asian Englishes. *English World-Wide* 40 (2): 115–143.
- Mair, Christian. 2009. Corpus linguistics meets sociolinguistics: The role of corpus evidence in the study of sociolinguistic variation and change. In A. Renouf and A. Kehoe (eds.), *Corpus linguistics: Refinements and reassessments*, 7–32. Leiden/Boston: Brill.
- Mauranen, Anna. 2011. Learners and users – Who do we want data from? In F. Meunier, S. Kock, G. Gilquin and M. Paquot (eds.), *A taste for corpora. In honour of Sylviane Granger*, 155–171. Amsterdam/Philadelphia: John Benjamins.
- Mauranen, Anna. 2018a. Second Language Acquisition, world Englishes, and English as a Lingua Franca (ELF). *World Englishes* 37 (1): 106–119.
- Mauranen, Anna. 2018b. Conceptualizing ELF. In J. Jenkins, W. Baker and M. Dewey (eds.), *The Routledge handbook of English as a lingua franca*, 7–24. London: Routledge.
- Milroy, James and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339–384.
- Osimk-Teasdale, Ruth. 2017. Analysing ELF variability. In J. Jenkins, W. Baker and M. Dewey (eds.), *The Routledge handbook of English as a lingua franca*, 201–209. London: Routledge.
- R Core Team. 2021. R: A language and environment for statistical computing, Vienna, Austria. Available at: <https://www.R-project.org/>.
- Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 2012. Exploring the interface between world Englishes and Second Language Acquisition – and implications for English as a Lingua Franca. *Journal of English as a Lingua Franca* 1 (1): 57–91.
- Seidlhofer, Barbara. 2004. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics* 24: 209–239.
- Seidlhofer, Barbara. 2013. *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Sharma, Devyani. 2017. World Englishes and sociolinguistic theory. In M. Filppula, J. Klemola and D. Sharma (eds.), *The Oxford handbook of World Englishes*, 232–251. Oxford: Oxford University Press.
- Smith, Nicholas. 2003. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In R. Facchinetti, M. Krug and F. Palmer (eds.), *Modality in contemporary English*, 241–267. Berlin/New York: De Gruyter.
- Suárez-Gómez, Cristina and Elena Seoane. 2023. The role of age and gender in grammatical variation in world Englishes. *World Englishes* 42: 327–343.
- Tagliamonte, Sali A. 2004. *Have to, gotta, must*. In C. Mair and H. Lindquist (eds.), *Corpus approaches to grammaticalization in English*, 33–55. John Benjamins.
- Tagliamonte, Sali A. and Alexandra D'Arcy. 2007. The modals of obligation/necessity in Canadian perspective. *English World-Wide* 28 (1): 47–87.
- Tagliamonte, Sali A. and R. Harald Baayen. 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.
- Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65 (1): 31–55.
- Traugott, Elizabeth Closs and Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

- Van Rooy, Bertus. 2010. Social and linguistic perspectives on variability in world Englishes. *World Englishes* 29 (1): 3–20.
- VOICE 2013. The Vienna-Oxford International Corpus of English (version 2.0 online). <https://voice2.acdh.oeaw.ac.at/index.xql> (last accessed 31 October 2023).
- Widdowson, Henry George. 2017. Historical perspectives on ELF. In J. Jenkins, W. Baker and M. Dewey (eds.), *The Routledge handbook of English as a lingua franca*, 101–112. London: Routledge.
- Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. London: Routledge.

Appendix 1. VIF scores for independent variables.

