CrossMark

*English Language and Linguistics*, 26.1: 133–159. © The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.
doi:10.1017/S1360674321000071

# Modal verbs of strong obligation in Scottish Standard English[1]

OLE SCHÜTZLER and JENNY HERZKY
*University of Bamberg*
(Received 20 April 2020; revised 19 February 2021)

This article investigates differences between Scottish Standard English (SSE) and Southern British Standard English (SBSE) in the semantic domain of strong obligation. Focusing on the modal verbs MUST, HAVE TO, NEED TO and (HAVE) GOT TO, we use new corpus material from nineteen written and spoken genres in the Scottish component of the *International Corpus of English* (ICE-SCO) and corresponding texts from ICE-GB. Data are analysed using a mixed-effect multinomial regression model to predict the choice of verb. Language-internal factors include mode of production (written/spoken), grammatical subject (first/second/third person) and source of obligation (objective/subjective). Our results show that, as previous research suggests, SSE is much more likely to employ NEED TO for the expression of strong obligation, and less likely to employ MUST and (HAVE) GOT TO. This general pattern remains essentially unaffected by language-internal factors. To account for our findings, we draw on the sociologically motivated process of democratisation and the language-internal process of grammaticalisation.

**Keywords:** modal verbs, strong obligation, grammaticalisation, democratisation, Scottish Standard English

## 1 Introduction

This article starts from the premise that not enough is known about grammatical differences between the standard Englishes of Scotland and England, here referred to as Scottish Standard English (SSE) and Southern British Standard English (SBSE), respectively.[2] In terms of its grammar, SSE is only theoretically recognised as a Standard variety, while empirical evidence concerning specific grammatical features is scarce. In our study, we focus on a central area of grammar, modal verbs of (strong) obligation.

[1] We gratefully acknowledge support from the DFG (German Research Council, grant SCHU 3250/1-1). We would also like to thank the editor and our two reviewers for their very helpful comments on an earlier version of this article.

[2] The language labels used are a constant source of headaches for linguists. Our convention suggests that they are regional/national variants of (World) Standard English (see McArthur 1987). 'Standard Southern British English' and 'Standard Scottish English' are alternative terms, but we caution against the use of 'Southern Standard British English' (as in McMahon 2002: 5, for example), which seems to presuppose some kind of general British Standard.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

Twentieth-century developments in the modal verb system of British and American English have collectively been dubbed 'modal decline' (e.g. Leech 2003, 2013; Smith 2003; Leech *et al*. 2009), which describes a decrease in the frequencies of core modals partly compensated by higher frequencies of semi-modals. Within the semantic domain of strong obligation, for example, the relative frequencies of the verbal predicates NEED TO and HAVE TO have increased at the expense of MUST (see Millar 2009: 204, 208–9; Leech 2013). For Scots (and Scottish English), Miller (2008: 305) claims that obligation is expressed by HAVE TO and NEED TO, while MUST is reserved for epistemic contexts. This and other evidence in the literature suggests that one difference between SSE and SBSE may lie in this domain.

Based on new corpus material from the Scottish component of the *International Corpus of English* (ICE-SCO) and corresponding texts from ICE-GB (representative of SBSE), we compare the relative frequencies of MUST, HAVE TO, NEED TO and (HAVE) GOT TO.[3] While the main focus is on intervarietal differences, we include mode of production (written vs spoken), grammatical subject (first, second and third person) and source of obligation (objective vs subjective) as predictors in a multivariate regression analysis.

Among other features, Schützler, Gut & Fuchs (2017) regard modal verbs as central to a better description of SSE: with the exception of (rare) double modal constructions, they are not overtly 'dialectal' – their characteristic behaviour in specific varieties like SSE is probabilistic and must be described in terms of frequencies, rather than categorical divergence. This view of intervarietal variation is of course commonplace in comparisons of British English (BrE) and American English (AmE) – see, for instance, contributions in Rohdenburg & Schlüter (2009). However, as we will argue in section 2, linguistic perspectives on SSE have for historico-political reasons been characterised by what Schützler, Gut & Fuchs (2017) call the 'Scots bias', i.e. the oversight of such probabilistic differences and a focus on more obvious features directly motivated from Scots. Our investigation aims to make first steps in overcoming this situation. The agenda of our article is twofold: (i) we present corpus-based evidence for a better comparison of SBSE and SSE, increasing the visibility of the latter as a variety, and (ii) we contribute methodologies and results to the existing research on English modal verb constructions more generally.

Section 2 is an introduction to SSE and some of the relevant research gaps. Section 3 summarises essential theoretical and empirical research background. In section 4, we formulate our research questions and hypotheses and explain our corpus-linguistic and statistical methodologies. The presentation of results in section 5 is followed by a conclusion and outlook in section 6.

[3] Modal NEED – as in '[Y]ou only need look over the border […]' (ICE-GB-S2A-036) – has been excluded due to the extremely low frequency of valid occurrences in both corpora. The vast majority of hits were negated forms, which were generally excluded (see section 4.2).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

## 2 Scottish Standard English, the 'Scots bias' and ICE-Scotland

*Scottish English* comprises a continuum ranging from Broad Scots (the vernacular) to SSE (McArthur 1979: 59; cf. Aitken 1979: 85). SSE is in principle recognised as a standard variety, but its phonological features tend to be highlighted relative to its grammatical characteristics, as in McClure (1994: 79):

> [SSE] is now an autonomous speech form, having the status of one among the many forms of the international English language, and is recognised as an established national standard, throughout the English-speaking world . . . Like other national forms of English, it is characterised to some extent by grammar, vocabulary and idiom, but most obviously by pronunciation.

In keeping with McArthur's (1987; 1998) notion of a World Standard English, it is natural to expect grammatical differences between standard dialects to be less salient compared to accent features. Similarly, Giegerich (1992: 45–6) associates distinctive lexical and grammatical features of Scottish English with Scots, perhaps thinking more in terms of categorical differences, i.e. constructions that exist in SSE but not in SBSE.

McClure (1994: 85) points out that phonetic and phonological characteristics of SSE are obvious, but that too little empirical research has been carried out on other linguistic levels (cf. exceptions discussed in Schützler, Gut & Fuchs 2017). A similar 'complaint' was made much earlier by Aitken (1979: 110), and, we would venture, can still be upheld today: beliefs and intuitions concerning the grammatical distinctness of SSE are supported by little empirical evidence. The question framed by McArthur (1979: 57) has largely gone unanswered, and can serve as the starting point for our research: 'What is the relationship between Scottish Standard English and the other national standards, and … how does it relate to its neighbour in southern England …?'

Research on the grammar of Scottish Englishes tends to focus on grammatical and/or lexical phenomena that are part of the inventory of Scots features (hence the 'Scots bias' discussed below; cf. Schützler, Gut & Fuchs 2017). This may partly reflect a subconscious bias: Scotland and England are not only directly adjacent but form a political union. This, in combination with the historically strong position of the Scots language, weakens the position of SSE as a standard.

Schützler, Gut & Fuchs (2017: 279) identify a number of interconnected 'deficits and gaps in research on SSE': (i) little interest particularly in grammatical features; (ii) a lack of suitable tools and data resources (e.g. corpora); (iii) a lack of research targeting SSE, which leads to (iv) a descriptive deficit regarding SSE in itself; and (v) difficulties in comparing SSE to other standard varieties of English. A beginning has been made in addressing the first two points by raising awareness of a 'Scots bias' which needs to be overcome, and by compiling the Scottish component of the *International Corpus of English* (ICE-Scotland).[4]

[4] For technical details concerning ICE-Scotland, see Schützler, Gut & Fuchs (2017). For a discussion of other corpus-linguistic resources for the study of Scottish Englishes, see Corbett & Stuart-Smith (2012) and Anderson (2013).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

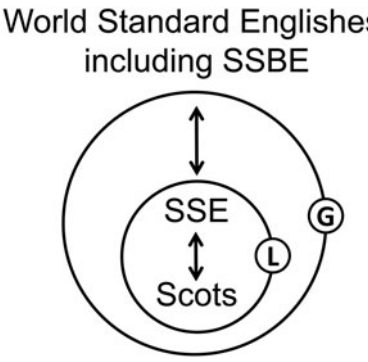


Figure 1. Local (L) and Global (G) continua of variation for SSE

Our view of SSE is shown in figure 1, which expands the model by McArthur (1979: 59; cf. Aitken 1979: 85) and is a slightly more compact version of the model developed in Schützler, Gut & Fuchs (2017: 282; cf. Schützler 2015: 25). We are interested in the standard pole (SSE) within the local continuum of variation and its position in an intervarietal context of variation – particularly concerning the influential close neighbour variety SBSE.

Since the grammar of SSE is not codified, it is challenging to decide what to treat as standard usage. Our approach is guided by the data source: since ICE is dominated by language in standard genres, produced by educated speakers and writers, the forms that are produced qualify as SSE.

In what follows, we will look into relevant theoretical aspects of the core grammatical category of modality

## 3 Modality and modal verbs of strong obligation

*Modality* in the widest sense is about changes made to the meaning of clauses to express assessments of truth values or factuality (Quirk *et al.* 1985: 221). Propositions are graded as more or less likely to be (or to become) factual, based on notions like possibility, permission, volition or obligation. Modality can be expressed in various ways (cf. Huddleston & Pullum *et al.* 2002: 52; Leech 2013: 109), but modal and semi-modal verbs receive the greatest attention in linguistic research.

Quirk *et al.* (1985: 219–20) distinguish two basic types of modality, (i) *intrinsic*, i.e. 'intrinsic human control over events' (permission, obligation, volition), and (ii) *extrinsic*, i.e. 'human judgement of likelihoods' (possibility, necessity, prediction). The extrinsic type is further subdivided, depending on whether or not it involves 'human control'. In the following examples from Quirk *et al.* (1985: 224–5), (1) is thus categorised as '(logical) necessity', which Quirk *et al.* equate with 'epistemic'; (2) is categorised as 'root necessity'; and (3) is classified as 'obligation or compulsion' and stands apart from (1) and (2).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

(1) There **must** be some mistake.
(2) To be healthy, a plant **must** receive a good supply of both sunshine and moisture.
(3) You **must** be back by ten o'clock.

Somewhat confusingly, there are two 'root' meanings of verbs like MUST in Quirk *et al.*'s scheme – exemplified as (2) and (3) – that belong to two different higher-level domains (necessity and obligation), although they seem closely related. For Huddleston & Pullum *et al.* (2002: 52), example (1) would be a case of epistemic modality, while (2) and (3) would both be deontic, involving an element of obligation. The difference between the latter two is captured by the dichotomy *subjective* vs *objective* (Huddleston & Pullum 2002 *et al.*: 183). In subjective deontic modality, the deontic source – i.e. '[t]he person, authority, convention, or whatever from whom the obligation, etc., is understood to emanate' (Huddleston & Pullum *et al.* 2002: 178) – coincides with the speaker/writer, or is clearly stated. In objective deontic modality, the speaker/writer states an obligation imposed by some other, more abstract agency. Thus, (4) and (5) are classified as subjective and objective, respectively (Huddleston & Pullum *et al.* 2002: 183).[5]

(4) You **must** clean up this mess at once.
(5) We **must** make an appointment if we want to see the Dean.

A third type of modality identified by Huddleston & Pullum *et al.* is *dynamic modality* (2002: 52, 185). Here, the obligation derives from personal (and perhaps situational) properties, as in (6) and (7) from Huddleston & Pullum *et al.* (2002: 185):

(6) Ed's a guy who **must** always be poking his nose into other people's business.
(7) Now that she has lost her job she **must** live extremely frugally.

Huddleston & Pullum *et al.* regard (6) as more prototypically dynamic: the necessity truly arises from the subject himself, while in (7) there is a more abstract 'force of circumstance'. The central criterion for dynamic modality is the absence of a deontic source outside the closely delimited subject (e.g. 'Ed'), or the situation itself. We would agree that in (6), the source of the obligation cannot be located outside Ed's character. In cases like (7), however, we would argue that circumstances, however abstract, may impose obligations and therefore constitute deontic sources.

Finally, Biber *et al.* (1999: 485) apply a binary categorisation of modal meaning into (i) *intrinsic* (or *deontic*), in which case the situation is controlled via some agent (human or other), and (ii) *extrinsic* (or *epistemic*), which refers to the likelihood status of events or states (see Sweetser 1990: 49). Crucially, Biber *et al.* (1999) do not rely on a human source of obligation for the deontic category. In our analysis, we will adopt this more general framework: clear cases of epistemic and dynamic modality (examples (1) and (6)) are excluded, and the remaining deontic cases are classified as subjective or objective.

[5] Ultimately, the deontic source in objective cases like (5) may also be human, but not transparently so.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

### 3.1 *Differences between modal verbs of strong obligation*

The verbs treated in this article belong in the category of strong obligation (see Huddleston & Pullum *et al.* 2002: 176–7 for a differentiation between weak, medium and strong modality), since the proposed action appears (nearly) compulsory (see Sweetser 1990: 54). We do not differentiate between stronger and weaker meanings between verbs, since these are 'virtually impossible to categorize impartially' (Tagliamonte & Smith 2006: 345).

In Quirk *et al.* (1985: 145), the four verbs are generally equivalent: HAVE TO and MUST are similar in meaning, HAVE TO and (HAVE) GOT TO are 'semantically parallel', and the relationship between all four is described as 'close' (226). More weight is given to grammatical differences. For instance, HAVE TO 'can stand in for MUST in past constructions where MUST cannot occur', and in contexts that require a nonfinite form. Such contexts also restrict the occurrence of (HAVE) GOT TO as opposed to HAVE TO (Quirk *et al.* 1985: 145):

(8) (a) I have to wait. / I (have) got to wait. / I must wait.
(b) I had to wait. / *I had got to wait. / *I musted wait.
(c) I will have to wait. / *I will (have) got to wait. / *I will must wait.
(d) I am having to wait. / *I am having got to wait. / *I am musting wait.

There is grammatical near-equivalence of NEED TO and HAVE TO: substituting the former for the latter is unproblematic in sentences (8a–c), but marginally unacceptable in (8d).

The literature further reports that different verbs tend to be selected depending on whether the deontic source is subjective or objective. Huddleston & Pullum *et al.* (2002: 183) find that objective cases are realised with HAVE (GOT) TO or NEED TO, rather than MUST (see also Smith 2003: 243–4), and Quirk *et al.* (1985: 225–6) see MUST more strongly associated with subjective obligation than the 'more impersonal' HAVE (GOT) TO. Sweetser (1990: 53) says that, compared to MUST, HAVE TO 'has more of a meaning of being obliged by extrinsically imposed authority' and NEED TO implies 'that the obligation is imposed by something internal to the doer' (cf. Talmy 1988). The view of NEED TO as expressing an internally motivated obligation for the agent's own sake is also found in Smith (2003: 244–5); we will discuss possible theoretical consequences of this in section 3.4.

### 3.2 *L1 varieties of English other than Scottish English*

This section summarises research on modal verbs of strong obligation in L1 varieties of English – mostly American and British English (hereafter: AmE and BrE). The literature-based discussion of Scottish Englishes follows in the next section.

While the studies that are summarised show us the general frequencies of our verbs, they often do not disambiguate deontic and epistemic meanings, and furthermore take a traditional approach in counting and normalising frequencies based on super-categories (e.g. whole corpora, or corpus sections) without considering the hierarchical structure of such data sources. Our approach (see section 4.3) thus limits

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

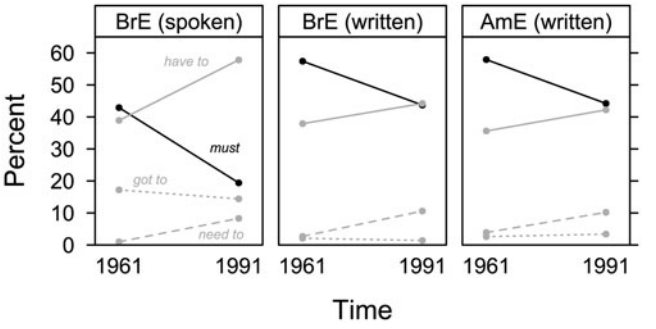


Figure 2. Modals of strong obligation in late twentieth-century BrE and AmE (based on Smith 2003: 248–9)

direct comparability with earlier studies. Finally, most studies report absolute (normalised) text frequencies, while we convert these to percentages relative to a category comprised of all occurrences of MUST, HAVE TO, NEED TO and (HAVE) GOT TO.[6]

Figure 2 summarises Smith's (2003: 248–9) analysis of BrE and AmE.[7] It is based on written data from four corpora in the Brown family (e.g. Francis & Kučera 1979), complemented by two smaller spoken corpora of BrE.[8] No spoken AmE data were included.

In the spoken BrE data, HAVE TO and (HAVE) GOT TO are generally more frequent than in writing; relative frequencies of MUST decrease and relative frequencies of HAVE TO and NEED TO increase over time. Since Smith looks at individual and cumulated text frequencies, some of his conclusions are not borne out by our representation, for example concerning differences between written BrE and AmE (which are virtually non-existent in figure 2).

In figure 3, we reproduce spoken data from the demographic (spoken) subcorpus of the *British National Corpus* and the *Longman Corpus of Spoken American English*, both representative of 1990s usage, as published in Leech (2013: 112). This analysis, too, is entirely form-driven.

Results in BrE agree with Smith's (2003) data in the first panel of figure 2, except that in the BNC (HAVE) GOT TO is more frequent. There is also a marked difference between BrE and AmE, the latter showing a more pronounced preference for HAVE TO and NEED TO, at the expense of MUST and (HAVE) GOT TO.

[6] In all our plots, we label the category (HAVE) GOT TO more simply as *got to*.

[7] As one reviewer pointed out, line plots should only be used if both the x-axis and the y-axis are interval-scaled. We agree that the lines in our plots may be interpreted as indicating intermediate values where none exist; we would also argue, however, that the advantages of line plots for the visualisation of complex data far outweigh such concerns (cf. Sönning 2018), which is why we decided to retain them.

[8] Earlier and later spoken BrE is represented by corpora called 'SEU mini' and 'ICE-GB mini', respectively; earlier and later written BrE is represented by the LOB and FLOB corpora; and earlier and later written AmE is represented by the Brown and Frown corpora. For more information see Smith (2003).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

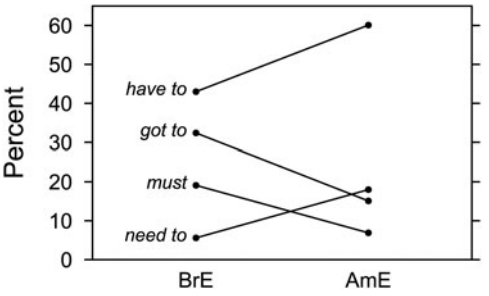


Figure 3. Modal verbs of strong obligation in spoken BrE and AmE (based on Leech 2013: 112)

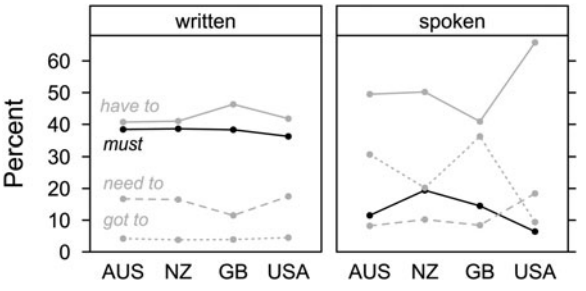


Figure 4. Modal verbs of strong obligation in L1 varieties of English (based on Collins 2009: 285–6)

In figure 4, we re-visualise central results from Collins (2009).[9] Our focus is on L1 varieties only. In writing, the pattern is relatively stable across the four varieties under investigation (Australian, New Zealand, British and US-American English), with high frequencies of HAVE TO and MUST, lower frequencies of NEED TO, and low frequencies of (HAVE) GOT TO – see results from Smith (2003) in figure 2. In speech, with few exeptions, HAVE TO and (HAVE) GOT TO are used more, while MUST is much less frequent and NEED TO is somewhat less frequent.

Finally, Millar (2009) inspects diachronic developments in the frequencies of modal verbs in the *TIME Magazine Corpus* (Davies 2007–). He finds a clear twentieth-century trend with MUST decreasing and HAVE TO and NEED TO increasing in frequency, while (HAVE) GOT TO does not change much.

### *3.3 Scottish English*

Differences between modal verb usage in SSE and SBSE are commented on by Quirk *et al.* (1985: 220), who say that 'Scots, Irish, and Northern English varieties resemble AmE in some respects more than they resemble the "standard" southern usage …'. In

[9] With the exception of AmE, Collins' analysis is based on ICE-corpora, which contain texts from a range of spoken and written genres assumed to be reasonably representative of standard usage (see appendix A for the genres we used in this article). For the AmE data, Collins' analysis draws on a combination of the (spoken) *Santa Barbara Corpus* and the *Freiburg–Brown Corpus of Written American English* (Frown).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

the following paragraphs, we avoid the issue of double-modal constructions – which may be what Quirk *et al.* allude to – although these have received a good deal of attention (e.g. Miller & Brown 1982: 12–13; Brown 1991; Fennel & Butters 1996: 273). In our data, such constructions did not occur, and it seems likely that double modals do not play a role in standard usage in Scotland.[10] The studies cited in this section do not explicitly target standard language, and it is therefore uncertain to what extent the described tendencies can be generalised to SSE.

Miller & Brown (1982: 8) list five modal verbs for the expression of necessity in Scottish English: MUST, HAVE TO, WILL HAVE TO, (HAVE) GOT TO and NEED TO. There is a strong restriction of MUST to its epistemic sense, as in (9) (see also Aitken 1979: 105; Kirk 1987; Miller 1989: 16–17; 2008: 305). The epistemic sense is generally regarded as more recent than the deontic sense, illustrated in (10) (cf. Sweetser 1990; Tagliamonte & Smith 2006):

(9) I keep thinking about the other tenants in the building and how they **must** be feeling. (ICE-SCO-rep-056)

(10) Scotland's mountains and wild lands are one of our greatest treasures and **must** be protected. (ICE-SCO-rep-062)

According to Miller & Brown (1982), it is NEED TO and HAVE TO that are mainly used for the expression of necessity. MUST is stronger (or more emphatic), but Miller & Brown suspect that there is no difference between HAVE TO and MUST concerning the deontic source. In the same context, NEED TO is described as 'no less strong than MUST or HAVE TO'. Finally, Miller & Brown (1982: 10) comment that the form WILL HAVE TO may be used where other varieties use MUST. Two examples from ICE-Scotland show that seemingly futurate uses of NEED TO and HAVE TO can be used with a meaning equivalent to a present-tense form:

(11) This candidate drug **will need to** undergo further preclinical testing before it can be taken forward into clinical trials […]. (ICE-SCO-PNat-13)

(12) [This] means that the government has adopted Labour's shale gas policy and **will have to** bring in new environmental regulations before fracking can be allowed. (ICE-SCO-rep-73)

Forms like these are perhaps only formally marked for future, without any future meaning proper.

Two Scottish dialects are included in Tagliamonte & Smith's (2006) study: Buckie (in the present-day council area of Moray) and Cumnock (East Ayrshire). For methodological reasons, their results cannot be straightforwardly related to our own findings. What can be said, however, is that the two regional Scottish varieties seem to be rather different (see Tagliamonte & Smith 2006: 366, table 3). Relative to the average behaviour of dialects, Buckie speakers simply do not use MUST, strongly prefer HAVE TO and disprefer (HAVE) GOT TO; Cumnock speakers, however, are very close indeed to the cross-dialectal

[10] Smith (2012) gives double modals in Scottish English the rating 'C' ('exists, but extremely rare') for *The Mouton World Atlas of Variation in English* (Kortmann & Lunkenheimer 2012).

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

average. The verb NEED TO is not included, which additionally complicates a direct comparison to our data. Using a selection of crossed language-internal factors, Tagliamonte & Smith (2006: 363) further find that a subjective deontic source correlates with MUST, and an objective one with HAVE TO, at least with first- and third-person subjects.

The main task of the analyses presented in section 5 will be to establish whether the (more vernacular) patterns reported above have a parallel in SSE as documented in ICE-Scotland.

### 3.4 *Mechanisms of change: grammaticalisation, democratisation and reallocation*

Like Tagliamonte & Smith (2006: 344), we interpret the coexistence of semantically equivalent modals or semi-modals as a case of layering in the sense of Hopper (1991: 22–4). In grammaticalisation (Hopper & Traugott 2003), existing forms take on new (grammatical or semantic) functions and become layered with older, functionally similar forms. For MUST, HAVE TO and HAVE GOT TO, the process is summarised by Tagliamonte & Smith (2006: 346–8): Germanic predecessors of MUST were part of Old English, while HAVE TO is first attested in Middle English, became more established in Early Modern English, and developed further – e.g. into (HAVE) GOT TO and its even more compacted and modalised form GOTTA – from the nineteenth century onwards (see also Krug 2000). NEED TO followed a grammaticalisation path roughly comparable to HAVE TO (*OED* online, *s.v.* 'need, *v.2*').

However, grammaticalisation as an intra-systemic mechanism of change does not explain *why* a change happens. There seem to be mainly two factors that can motivate changes in our verbs. The first one assumes a process of *democratisation* (e.g. Farrelly & Seoane 2012), which Leech *et al.* (2009: 259) define as a discourse-pragmatic (i.e. linguistic) correlate of 'changing norms in personal relations' – a kind of language change directly linked to (and caused by) changes in society. The adoption of different modal verbs for the coding of strong obligation would seem to fall into a category that Schützler (2020) calls 'explicit democratisation', i.e. the avoidance of overt linguistic markers of inequality or power asymmetry (see Fairclough 1992). This is consistent with Smith's (2003: 263) description of deontic MUST as 'prototypically subjective and insistent, sometimes authoritarian-sounding' and therefore 'likely to be increasingly avoided in a culture where overt markers of power or hierarchy are much less in favour'. The avoidance of MUST can then be compensated for by increased frequencies of HAVE TO and NEED TO. Associated with objective deontic sources, the former avoids the impression of a direct imposition of speaker authority. Concerning NEED TO, Smith (2003: 263–4; cf. Leech 2003: 237) writes that it 'is (ostensibly at least) not an overt marker of power' and can thus serve as 'an indirect means of laying down obligations'. As Smith (2003: 244–5) points out, NEED TO may express a (potentially strong) directive that poses, so to speak, as a recommendation in the addressee's own interest.

Secondly, and notwithstanding possible differences in their typical deontic source configuration (see section 3.1), the basic functional equivalence of verbs postulated in Quirk *et al.* (1985; cf. Talmy 1988: 86) makes it quite likely that they have been

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

reallocated to socio-stylistic (including dialect-marking) functions (see Trudgill 1986: 118–21; Britain & Trudgill 2000: 73–4). Stylistic differences of the verbs are discussed by Talmy (1988: 77), who comments on the colloquial character of HAVE TO relative to MUST, Leech *et al.* (2009: 95), who remark upon (HAVE) GOT TO as a colloquialism, as well as Biber *et al.* (1999: 489), who compare a number of verbs concerning their distribution across broad genres. We will make such comparison at a very general level by comparing speech and writing, since our central concern is the function of verbs as dialect markers that signal a difference between SBSE and SSE.

## 4 Research questions, data and methodology

### *4.1 Research questions and expectations*

Our main research question is whether or not SSE does indeed follow different strategies when encoding strong obligation with the verbs MUST, HAVE TO, NEED TO and (HAVE) GOT TO. Miller (1989: 17) foreshadows our expectation:

> My impression … is that NEED occurs frequently. I at one time imagined that it was more frequent than HAVE TO, but this is not borne out by the recorded data. It may well be, however, that NEED is used more frequently by speakers of Scottish English than by speakers of other varieties.

We expect that MUST is used less in SSE, while HAVE TO and particularly NEED TO are used more, compared to SBSE. We further expect those differences to surface more strongly in spoken language, since writing will be characterised to a greater extent by the kind of 'text-linked World Standard' postulated by McArthur (1987: 10).

Due to the socio-stylistic values of verbs, we expect higher relative frequencies of HAVE TO, NEED TO and (HAVE) GOT TO in speech, while MUST retains a relatively central position in writing. Further, if we accept the association of MUST with overtly expressed authority, this verb should be less frequent in connection with subjects in the second-person. Finally, we expect an association of HAVE TO with objective obligation. Apart from general differences between varieties, we will also inspect whether or not intra-linguistic or contextual factors (grammatical subject, deontic source, mode of production) have similar effects in SSE and SBSE, or whether they, too, are variety-specific.

### *4.2 Data retrieval and coding*

We used two components of the *International Corpus of English* (ICE): ICE-GB (Nelson, Wallis & Aarts 2002; Kirk & Nelson 2018) and ICE-Scotland ('ICE-SCO'; Schützler, Gut & Fuchs 2017). We included material from 21 genres. With the concordancing software *AntConc* (Anthony 2018), we retrieved occurrences of <must>, <have to>, <has to>, <need to>, <needs to> and <got to>. Nonfinite forms (e.g. *will have to*, *might need to*), interrogatives, past-tense forms, as well as negations were excluded, as were all epistemic instances (predominantly of MUST) and non-obligation meanings of *got to* (e.g.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

Table 1. *Raw counts and percentages of verbs*

| | | MUST | HAVE TO | NEED TO | (HAVE) GOT TO |
|---|---|---|---|---|---|
| SBSE (ICE-GB) | *n* | 240 | 285 | 84 | 50 |
| | % | 36 | 43 | 13 | 8 |
| SSE (ICE-SCO) | *n* | 65 | 79 | 89 | 6 |
| | % | 27 | 33 | 37 | 3 |

*They got to be friends*). The categories 'legal cross-examinations' and 'business letters' – at the time of analysis represented by $n = 4$ and $n = 6$ texts, respectively, in ICE-SCO – did not yield valid hits in SSE and were excluded to maintain the genre balance between corpora. Eventually, the analysis was thus based on $n = 19$ genres (9 spoken, 10 written). Genres and total numbers of texts and words are documented in appendix A.

We obtained $n = 898$ tokens, distributed across $n = 332$ individual texts. This number of texts is considerably lower than the number that were searched ($n = 607$) – not necessarily because none of our verbs occurred in the remaining $n = 275$ texts, but because the retrieved cases did not meet the necessary grammatical and semantic criteria. Raw counts and percentages (per variety) are shown in table 1.

Cases were coded for the variables shown in table 2. Grammatical coding and the exclusion of false positives and epistemic cases were done by the second author, supported by Zeyu Li at the University of Münster. Rare instances of disagreement were discussed and resolved in cooperation with author one. Concerning deontic source, both authors independently awarded scores of −1 (objective), +1 (subjective) and 0 (intermediate/unclassified). Cases of disagreement were only resolved if there

Table 2. *Definition of variables*

| Variable | Class | Scale/levels | Comments |
|---|---|---|---|
| VERB | outcome | *must*<br>*have to*<br>*need to*<br>*(have) got to* | Reference category: *must* |
| SCOT | fixed | ENG, SCO | Reference category: ENG (= SBSE) |
| SPOKEN | fixed | −1, +1 | written: −1, spoken: +1 |
| SUBJECT | fixed | *first*<br>*second*<br>*third* | Reference category: first (= 1st person) |
| SOURCE | fixed | −1, −0.5, 0, +0.5, +1 | unambiguously objective: −1; unambiguously subjective: +1 |
| GENRE | random | 19 levels | |
| TEXT | random | 332 levels | |

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

Table 3. *Deontic source: agreement between raters*

| | | Category author 2 | | |
|---|---|---|---|---|
| | | −1 | 0 | +1 |
| Category author 1 | −1 | **334** | 10 | 85 |
| | 0 | 20 | **4** | 26 |
| | +1 | 79 | 6 | **334** |

had been an obvious error. In other cases, the two ratings were averaged – thus, the predictor SOURCE can take five numerical values.

Our way of handling semantic classification reduces loss of information as well as the pressure involved in making a forced decision. The two sets of ratings are cross-tabulated in table 3.

Relative observed agreement was at 74.8 per cent: both raters awarded the same score in 672 out of 898 cases. Considering that we are dealing with semantic disambiguation, this is a good rate, even if complete disagreement (with exactly opposite ratings) was also substantial (164/898 = 18.3 per cent). Apart from true inter-rater errors, this figure may partly be explained from the fact that (i) it was not feasible to inspect the full context of examples, and that (ii) many examples are truly ambiguous. By averaging the two scores for deontic source, we neutralise conflicting cases in a consistent, non-lossy way.

### *4.3 Statistical modelling and visualisation*

We worked in the R-environment (R Core Team 2019), using *RStudio* (RStudio Team 2009–19). Visualisations were done with functions in the R-package 'lattice' (Sarkar 2018). With the variables in table 2, we fitted a Bayesian multinomial mixed-effects model to the data, using the R-package 'brms' (Bürkner 2020), which is based on Stan (Stan Development Team 2019). A multinomial model has more than two possible categorical outcomes, whose probabilities under the influence of different factors are estimated. In this case there are four possible outcomes corresponding to the four modal verbs under investigation, encoded in the variable VERB. Predictor variables were SUBJECT, SOURCE and SPOKEN, each of which was specified as interacting with variety. In other words: the model allowed for the possibility that SUBJECT, SOURCE and SPOKEN take different effects on the selection of verbs in SBSE and SSE. For the exact model specification, a discussion of the priors that were used, as well as for further information see appendix B.

The plots show median values of the estimated percentages of the four verbs under different conditions, as well as their dispersion, expressed as 50% and 90% percentile-based posterior uncertainty intervals. Such intervals will sometimes be reported in the text. For instance, in figure 5, the percentage point difference between spoken SSE and spoken SBSE for NEED TO can be given as '22.1% [10.3, 36.3]', which

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

means 'a median difference of 22.1% with a 90% uncertainty interval extending from 10.3% to 36.3%'.

Factors of no immediate interest in a given scenario are held constant. The exact routine is made transparent on https://osf.io/aq2r5/ (see section 4.4 below). Intuitively, it may seem to be problematic to assume normal (or average) values for mode of production or grammatical subject, since, in reality, these parameters take categorical values at any one time.[11] However, this approach allows us to target specific effects and thus makes results more accessible.

Concerning the fixed part, no model comparison or model selection process was conducted, since it was considered essential to retain all theoretically important predictors in the model, irrespective of their estimated effects (cf. Heinze, Wallisch & Dunkler 2018); we thus adopted the notion of the 'deductive model' as proposed by Tizón-Couto & Lorenz (2015). No *p*-values are calculated for individual coefficients, since we prefer an estimation approach to (less informative) null hypothesis testing. The reader can assess the robustness of an effect by looking at the uncertainty interval of the estimated difference (the effect size). If the interval cuts across the critical value of zero, results need to be treated with caution. The table in appendix C provides the complete fixed-effects part of the model output, as well as basic information on the random effects. Concerning the random part, the inclusion of TEXT as a cluster variable seemed absolutely necessary, and in our inclusion of GENRE we were guided by the structure of the ICE corpus. The random-effects structure at the level of TEXT is maximal, i.e. it mirrors the fixed effects and thus makes them more precise.

### *4.4 Open data*

The dataset used in the present study is published as Schützler & Herzky (2021) at the *Tromsø Repository of Language and Linguistics* (TROLLing; see references). R-scripts used in the analysis can be retrieved from https://osf.io/aq2r5/. Readers are thus enabled to understand our approach more fully, incorporate our data into their own analyses, adapt our models (e.g. by using different priors, including more interactions, or specifying different random effects) or to implement altogether different kinds of models (e.g. of a non-Bayesian type) or statistical tests. The osf-repository also contains the scripts for the generation of figures 5–10, along with the figures themselves (in svg-format). For the entire repository, we selected a CC BY 4.0 licence (https://creativecommons.org/licenses/by/4.0/) and marked the respective figures in the captions, thus: ⓒⓘ.

## 5 Results

Each of the following subsections focuses on one of three contrasts: speech vs writing, objective vs subjective sources of obligation, and the differences between grammatical

[11] Regarding deontic source, this is perhaps not equally the case if we allow for ambiguity; see discussion in section 4.2.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

subjects. For different conditions, estimated percentages of the four verbal categories are plotted for SSE and SBSE, controlling for other factors. Additionally, the percentage point differences between the two varieties are estimated and plotted. The reader can thus see at a glance (i) the expected proportions of verbal categories in each variety and (ii) the magnitude and robustness of the difference between SSE and SBSE. In addition, the effects of a difference (like speech vs writing) on specific verbs are gauged.

### *5.1 Speech and writing*

We turn first to the difference between modes of production. The bottom panels in parts (a) and (b) of figure 5 show estimated percentages of the four verbs in speech and writing; the panels on the top compare the two variety-based patterns by subtracting percentages in SBSE from percentages in SSE. For the sake of clarity, we use the labels 'SCO' and 'ENG', rather than 'SSE' and 'SBSE'.

In speech, the preferred verb in both varieties is HAVE TO, at 53.6% in SBSE and 49.6% in SSE. The difference is small enough to be passed over quickly. NEED TO is used at rates of 33.1% in SSE and 10.6% in SBSE, respectively. This preference for NEED TO in SSE is one of the two main differences between varieties. The second major difference concerns the relative frequency of MUST, which is 23.6% in SBSE and 9.5% in SSE. Again, the difference is quite robust. Finally, (HAVE) GOT TO is used more frequently in SBSE (12.7%) than in SSE (6.0%).

In writing, the general pattern is fundamentally different. In both varieties, the percentage of MUST is higher than in speech, with 38.4% in SBSE and 19.9% in SSE; percentages of HAVE TO are lower but still fairly similar in both varieties; frequencies of NEED TO are somewhat higher; and frequencies of (HAVE) GOT TO are lower than in speech. Thus, all verbs, perhaps with the exception of NEED TO, respond to the difference between modes. Crucially, however, the systematic differences between SSE

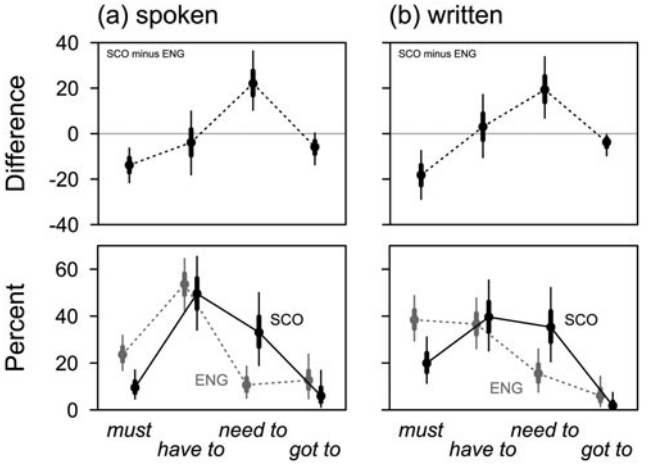


Figure 5. SSE vs SBSE: percentages and differences by verb and mode of production

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

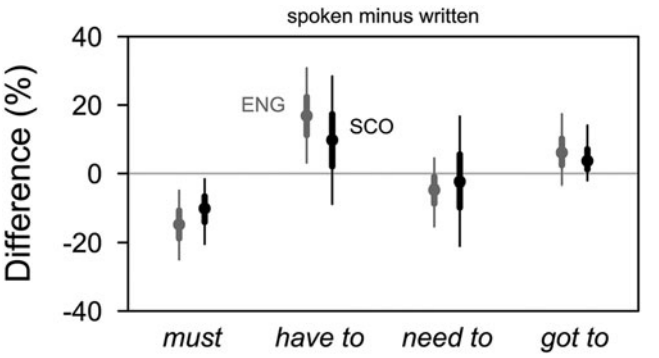


Figure 6. Speech vs writing: differences by verb and variety

and SBSE hold across speech and writing, as shown in the top panels of figure 5: MUST and (HAVE) GOT TO are more frequent in SBSE, NEED TO is more frequent in SSE, and HAVE TO is roughly equally frequent. The following two examples show an instance of written MUST and spoken HAVE TO from ICE-SCO.

(13) Current legislation dictates that sprinklers **must** be fitted in all care homes and sheltered housing […] (ICE-SCO-rep-056)

(14) […] this is a scandal […] we **have to** expose what went on here in Scotland […] (ICE-SCO-btal-036)

Figure 6 reorganises the information contained in figure 5 to focus on how the frequencies of individual verbs are affected by modes of production. Individual verbs' responses to the two conditions are more clearly visible here. Two verbs – HAVE TO and and (HAVE) GOT TO – associate with speech, MUST associates with writing, and NEED TO is relatively indifferent in SSE, perhaps tending towards the written mode in SBSE. Crucially, the effects point in the same direction in both varieties, although they differ in magnitude. Whatever the general differences in patterns between SBSE and SSE, the stylistic value of the four verbs seems to be similar, at least at this general level.

In section 3.4, we discussed the more colloquial character of HAVE TO and (HAVE) GOT TO. If we accept speech and writing as very broad stylistic categories, the behaviour of the two verbs in our data is consonant with those accounts.

### *5.2 Sources of obligation*

In this section, we inspect subjective and objective sources of obligation, holding other factors constant. As discussed in section 3.1, the source of objective authority lies outside the speaker or writer, which is often the case in rules and regulations, while in subjective cases the authority is imposed by the speaking or writing subject (Huddleston & Pullum *et al.* 2002: 183; Tagliamonte & Smith 2006: 361–2).

Figure 7 shows estimated percentages of the four verbs in the bottom panels, and compares the two variety-based patterns in the top panels. Irrespective of the results we discuss below, the general difference between the two varieties holds true from this

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

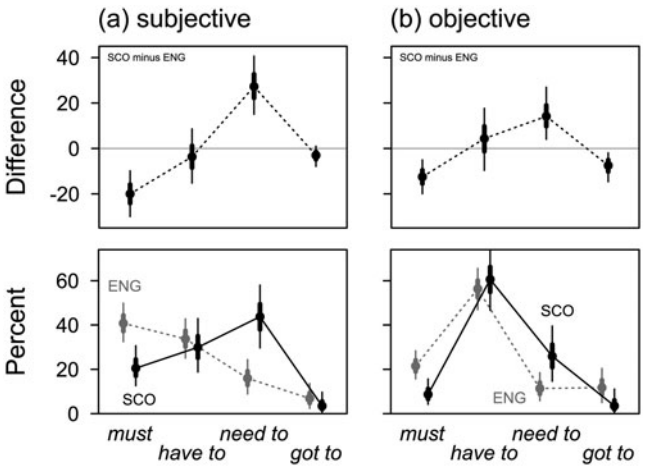


Figure 7. SSE vs SBSE: percentages and differences by verb and source of obligation

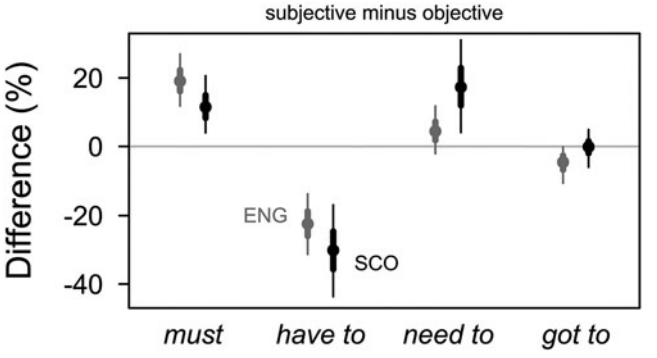


Figure 8. Subjective vs objective source of obligation: differences by verb and variety

perspective, too: Compared to SBSE, NEED TO is substantially more frequent in SSE, MUST is substantially less frequent, and (HAVE) GOT TO is somewhat less frequent.

Further, the difference between subjective and objective obligation correlates with the choice of verb in similar ways as the difference between writing and speech. Due to this similarity of the effects, there is considerable similarity between figures 5a and 7b, as well as between figures 5b and 7a, and we can therefore let figure 7 speak for itself.

In analogy to figure 6, figure 8 focuses on the behaviour of verbs between conditions. Subjective obligation disfavours HAVE TO in both varieties (see Tagliamonte & Smith 2006: 362): Compared to objective obligation, the respective relative frequencies of HAVE TO are lower by 22.4% [13.7, 31.3] in SBSE and 30.1% [16.9, 43.7] in SSE.[12] On the other hand,

[12] Note that we report (absolute) percentage-point differences. Further, we report all differences as positive; the direction of the effect is made explicit in the text and in the plots.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

MUST is more common with this type of deontic source (see Huddleston & Pullum *et al.* 2002). In our data, NEED TO is associated with subjective sources of obligation, particularly in SSE, where its relative frequency is higher by 17.3% [4.1, 31.1] in this condition, compared to objective deontic sources.

We conclude that the general difference between SBSE and SSE concerning relative frequencies of MUST, NEED TO and (HAVE) GOT TO is robust, even if we isolate the two deontic sources. Secondly, the impact of deontic source on the choice of modal verb is similar in both varieties.

### *5.3 Grammatical subjects*

The analysis of the relationship between grammatical subjects and verb selection is guided by the idea that notionally authoritarian items like MUST should occur less frequently when there is a second-person addressee. Figure 9 shows frequency patterns in a form familiar from above, with the direct comparison of SSE and SBSE in the top panels. We concentrate on those characteristics that stand out in each condition.

With second-person subjects, MUST is considerably less frequent than with first- or third-person subjects, at 6.4% in SSE and 20.1% in SBSE. At the same time, HAVE TO is the majority variant in both varieties (SSE: 52.8%; SBSE: 49.9%). Third-person subjects, however, associate strongly with the verb MUST, which is the majority variant in both varieties in this condition (SSE: 32.6%; SBSE: 39.7%). This is remarkable, given that SSE generally uses this verb less when compared to SBSE (see above). As a result, the pattern recurrently observed in the top panels of earlier figures – less MUST and more NEED TO in SSE, relatively indifferent behaviour of HAVE TO and (HAVE) GOT TO – is qualified with third-person subjects: In this context, the otherwise marked difference between varieties concerning frequencies of MUST virtually disappears. Only a small and not particularly robust difference remains.

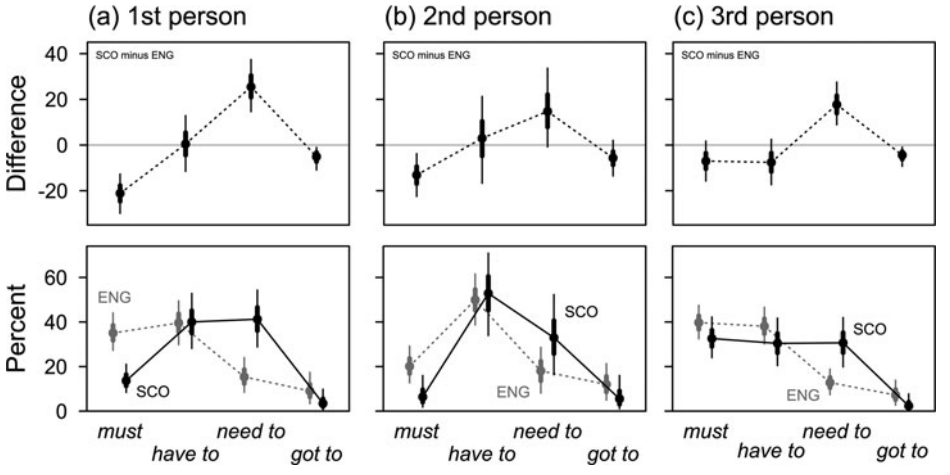


Figure 9. SSE vs SBSE: percentages and differences by verb and grammatical subject

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

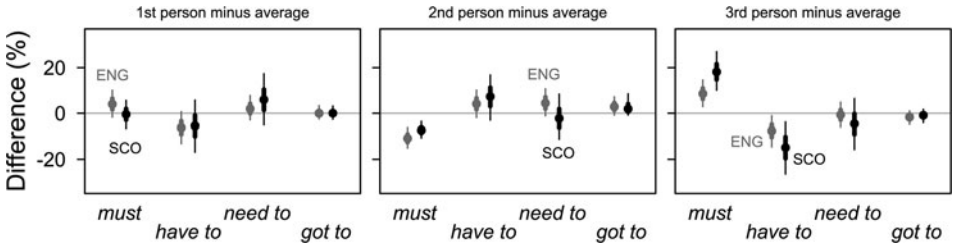


Figure 10. Grammatical subjects: differences by verb and variety

For the direct inspection of the behaviour of individual verbs in figure 10, we compare the estimated frequencies of verbs in combination with any of the three subject conditions to an idealised average value (see section 4.3). This was preferred to a less informative pairwise comparison of subject conditions.

Figure 10 highlights the main differences discussed above. Variation conditioned by first-person subjects is rather limited; with second-person subjects, percentages of MUST are much lower in both varieties, by 7.3% [3.5, 10.7] in SSE and by 10.8% [6.3, 15.1] in SBSE; in the third person, MUST is above average by 18.1% [10.1, 27.0] in SSE and by 8.6% [2.9, 14.6] in SBSE, while HAVE TO is below average by 14.8% [3.8, 26.5] in SSE and by 7.6% [1.1, 14.6] in SBSE. The verbs NEED TO and (HAVE) GOT TO do not respond much to grammatical-subject conditions. For illustration, the following two examples show third-person subjects in combination with MUST and HAVE TO, respectively.

(15) SNP **must** face economic realities (ICE-SCO-ed-021)
(16) [T]he Labour Party **has to** be at the forefront of the Labour Movement (ICE-SCO-nbtal-011)

A straightforward reading of the examples as conditioned by grammatical subjects is hampered not only by the headline style of (15), but also by the fact that deontic source seems to play a major role as well, probably with an objective source in (15) and a subjective one in (16).

## 6 Conclusion and outlook

There are both similarities and systematic differences between SSE and SBSE concerning the use of the four modal verbs of strong obligation, MUST, HAVE TO, NEED TO and (HAVE) GOT TO. Against the background of rather different baseline frequencies of the four verbs, the contextual, semantic and syntactic factors that govern the concrete choice of verb are remarkably similar:

i. In both varieties, MUST and – to a much lesser extent – NEED TO are associated with written language, while HAVE TO and – to a lesser extent – (HAVE) GOT TO are associated with spoken language. If we accept that the basic difference between speech and writing correlates with basic stylistic categories, our results are in agreement with descriptions of MUST as more formal than HAVE TO and (HAVE) GOT TO. The correlation of NEED TO with writing is a new insight.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

ii. In both varieties, subjective deontic sources of obligation correlate positively with higher rates of MUST and NEED TO, while objective deontic sources correlate with HAVE TO. For MUST, this pattern has been reported before, but to the best of our knowledge it has not been connected with NEED TO. The association of NEED TO with subjective deontic sources may be a yet unknown characteristic of its grammaticalisation path in English more generally. However, we regard it as an interesting indication rather than a conclusive result, and its further exploration must be left to independent follow-up research.
iii. In both SSE and SBSE, MUST tends to be used less with grammatical subjects in the second person, while HAVE TO is used more in this context. In contrast, with third-person subjects, there are increased frequencies of MUST and lower frequencies of HAVE TO. These patterns may be due to face-saving strategies: when directly telling someone what s/he should do, HAVE TO with its implied objective-obligation meaning is substituted for overtly authoritarian MUST.

Independently from the above constraints, both varieties appear to be characterised by different basic preferences concerning modal verbs of strong obligation. In SSE, NEED TO is used more frequently than in SBSE; in contrast, SBSE shows higher relative frequencies of (HAVE) GOT TO and particularly of MUST. We would argue that SSE has developed NEED TO as a strongly grammaticalised alternative to MUST, which is supported by the fact that both verbs are functionally similar, i.e. associated with writing and subjective deontic sources. Like HAVE TO, the verb NEED TO is 'softer' in a social sense, since it is traditionally linked to self-motivated obligation (cf. Talmy 1988; Sweetser 1990; Smith 2003). Although in our data NEED TO – like MUST – is linked to subjective (i.e. directly imposed) obligation and possibly to formal contexts, it may still serve as a more diplomatic (or democratic) functional equivalent of MUST.

With regard to our study, it is of course problematic that there is a considerable time gap between ICE-GB and ICE-SCO. As some of the research summarised in section 3 shows, there have been substantial diachronic developments in the system of modal verbs between the 1960s and the 1990s, and such trends may well have continued up to the present day. There is no easy solution to this problem: we have no corpora that represent SSE in the 1990s, and those corpora that could be used to assess present-day SBSE are either written-only – e.g. the BE06 corpus (Baker 2009) – or follow sampling frames different from ICE (e.g. BNC2014; Love *et al*. 2017). The way forward is likely to be some kind of corpus-triangulation approach, in which a more robust picture is assembled from carefully handled heterogeneous sources (ICE-SCO, BNC2014, Brown-family corpora). Given this problem, we make no claims for our research to be definitive; rather, we offer it as a reference point for future efforts, including our own.

We see three possible extensions of our present research. First, there is a distinction not only between grammatical subjects in different persons, but also between what has been called 'definite' vs 'indefinite' (or 'generic') subjects (see Tagliamonte & Smith 2006). The following examples illustrate the difference, showing definite subjective (17), definite objective (18) and indefinite objective cases (19), respectively:

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

(17) You **must** come and visit us as soon as you can. (ICE-GB-W1B-004)
(18) Yet the design **has to** be visible and distinctive so potential readers can quickly select their favourite newspaper […]. (ICE-SCO-ed-32)
(19) To work in high performance sport you **have to** be determined […] (ICE-SCO-Rep-09)

Including the definite-generic distinction in the analysis would certainly require more data. Diachronic and synchronic corpora large enough for this kind of undertaking exist for BrE and AmE, but it is doubtful whether ICE-SCO, even in its completed form, will be large enough.

Secondly, quasi-futurate forms – illustrated in examples (11) and (12) in section 3.3 – may play a special role in SSE. Future meaning is an implicit concomitant of obligation, as pointed out by Quirk *et al*. (1985: 217) with regard to MUST. In many constructions that combine finite WILL with the infinitive of HAVE TO or NEED TO, temporal deixis may therefore be much less important than a general softening effect achieved by making futurity explicit and thus reducing the immediacy of the imposed obligation. However, incorporating this into our quantitative approach would involve similar quantity-related issues as the inclusion of the definite-generic distinction of subjects. In our data, WILL HAVE TO and WILL NEED TO were rare, but they may well be one of the more subtle grammatical Scotticisms in SSE.

Despite these caveats and reservations, we have taken a step towards a better description of SSE concerning a central element of grammar. While, as expected, SSE and SBSE draw on the same basic system of modal verbs of obligation, the categories involved seem to have grammaticalised to different extents in the two varieties. SBSE is more traditional in using MUST at higher rates, but has also developed (HAVE) GOT TO as an alternative (see Krug 2000). In SSE, NEED TO has grammaticalised much more strongly, while (HAVE) GOT TO is less frequent. The notion of greater or lesser conservatism cannot be applied across the board. Rather, modal verbs grammaticalised along different trajectories in both varieties.

Concerning our object of investigation, one could perhaps speak of a British standard in the sense that both SSE and SBSE respond very similarly to factors like mode of production, deontic source and grammatical subject configuration. This shared set of rules, however, contrasts with different preferences of a more general kind. Consequently, the two major standard dialects in mainland Britain are characterised by unity and diversity at different levels. We will take this forward as a working hypothesis for our ongoing research on this topic.

*Authors' addresses:*

*Institute for English and American Studies*
*University of Bamberg*
*An der Universität 9*
*96047 Bamberg*
*Germany*
*ole.schuetzler@uni-bamberg.de*
*jenny.herzky@uni-bamberg.de*

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

## References


Aitken, Adam J. 1979. Scottish speech: A historical view with special reference to the Standard English of Scotland. In Adam J. Aitken & Tom McArthur (eds.), *Languages of Scotland*, 85–118. Edinburgh: Chambers.

Anderson, Wendy (ed.). 2013. *Language in Scotland: Corpus-based studies*. Amsterdam: Rodopi.

Anthony, Laurence. 2018. *AntConc: A freeware corpus analysis toolkit for concordancing and text analysis*, version 3.5.7 [computer program] www.laurenceanthony.net/software.html

Baker, Paul. 2009. The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics* 14(3), 312–37.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Britain, David & Peter Trudgill. 2000. Migration, dialect contact, new-dialect formation and reallocation. In Klaus J. Mattheier (ed.), *Dialect and migration in a changing Europe*, 73–8. Frankfurt: Peter Lang.

Brown, Keith. 1991. Double modals in Hawick Scots. In Peter Trudgill & Jack K. Chambers (eds.), *Dialects of English: Studies in grammatical variation*, 74–103. London: Longman.

Bürkner, Paul-Christian. 2020. *brms. Bayesian Regression Models using 'Stan'*. R-package version 2.12.0. https://cran.r-project.org/web/packages/brms/index.html

Collins, Peter. 2009. Modals and quasi-modals in world Englishes. *World Englishes* 28(3), 281–92.

Corbett, John & Jane Stuart-Smith. 2012. Standard English in Scotland. In Raymond Hickey (ed.), *Standards of English: Codified standards around the world*, 72–95. Cambridge: Cambridge University Press.

Davies, Mark. 2007–. *TIME Magazine Corpus: 100 million words, 1920s–2000s*. www.english-corpora.org/time/ [accessed 5 March 2020].

Facchinetti, Roberta, Manfred Krug & Frank Palmer (eds.). 2003. *Modality in contemporary English*. Berlin: Mouton de Gruyter.

Fairclough, Norman. 1992. *Discourse and social change*. Cambridge: Polity.

Farrelly, Michael & Elena Seoane. 2012. Democratization. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 392–401. Oxford: Oxford University Press.

Fennel, Barbara A. & Ronald R. Butters. 1996. Historical and contemporary distribution of double modals in English. In Edgar Schneider (ed.), *Focus on the USA*, 265–88. Amsterdam and Philadelphia: John Benjamins.

Francis, W. Nelson & Henry Kučera. 1979 [1971, 1964]. *A Standard Corpus of Present-Day Edited American English, for use with digital computers (Brown)*. Providence, RI: Brown University. http://korpus.uib.no/icame/manuals/FROWN/INDEX.HTM [accessed 4 March 2020].

Giegerich, Heinz J. 1992. *English phonology.* Cambridge: Cambridge University Press.

Heinze, Georg, Christine Wallisch & Daniela Dunkler. 2018. Variable selection: A review and recommendations for the practicing statistician. *Biometrical Journal* 60, 431–49.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 1: *Focus on theoretical and methodological issues*, 17–35. Amsterdam: John Benjamins.

Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Huddleston, Rodney D. & Geoffrey K. Pullum *et al*. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Kirk, John. 1987. The heteronomy of Scots with Standard English. In Caroline Macafee & Iseabail Macleod (eds.), *The Nuttis Schell: Essays on the Scots language presented to A. J. Aitken*, 166–81. Aberdeen: Aberdeen University Press.


https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

Kirk, John & Gerald Nelson. 2018. The International Corpus of English project: A progress report. *World Englishes* 34(4), 697–716.
Kortmann, Bernd & Kerstin Lunkenheimer (eds.). 2012. *The Mouton world atlas of variation in English*. Berlin: Mouton de Gruyter.
Krug, Manfred. 2000. *Emerging English modals: A corpus-based study of grammaticalization*. Berlin: Mouton de Gruyter.
Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In Facchinetti, Krug & Palmer (eds.), 223–40.
Leech, Geoffrey. 2013. Where have all the modals gone? An essay on the declining frequency of core modal auxiliaries in recent standard English. In Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), *English modality: Core, periphery and evidentiality*, 95–115. Berlin: Mouton de Gruyter.
Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.
Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3), 319–44.
McArthur, Tom. 1979. The status of English in and furth of Scotland. In Adam J. Aitken & Tom McArthur (eds.), *Languages of Scotland*, 50–67. Edinburgh: Chambers.
McArthur, Tom. 1987. The English languages? *English Today* 3(3), 9–11.
McArthur, Tom. 1998. *The English languages*. Cambridge: Cambridge University Press.
McClure, J. Derrick. 1994. English in Scotland. In Robert Burchfield (ed.), *The Cambridge history of the English language*, vol. 5: *English in Britain and overseas: Origins and development*, 23–93. Cambridge: Cambridge University Press.
McMahon, April M. S. 2002. *An introduction to English phonology*. Edinburgh: Edinburgh University Press.
Millar, Neil. 2009. Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics* 14(2), 191–220.
Miller, Jim E. 1989. The grammar of Scottish English. In James Milroy & Lesley Milroy (eds.), *Regional variation in British English syntax*. Swindon: Economic and Social Research Council.
Miller, Jim E. 2008. Scottish English: Morphology and syntax. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English*, vol. 1: *The British Isles*, 299–327. Berlin: Mouton de Gruyter.
Miller, Jim E. & Keith Brown. 1982. Aspects of Scottish English syntax. *English World-Wide* 3(1), 3–17.
Nelson, Gerald, Sean Wallis & Bas Aarts. 2002. *Exploring natural language: Working with the British component of the International Corpus of English*. Amsterdam and Philadelphia: John Benjamins.
*Oxford English Dictionary Online*. www.oed.com [accessed 6 March 2020].
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Arnold.
R Core Team. 2019. *R: A language and environment for statistical computing*, version 3.6.2. Vienna: R Foundation for Statistical Computing. www.r-project.org
Rohdenburg, Günter & Julia Schlüter (eds.). 2009. *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press.
RStudio Team. 2009–19. *RStudio: Integrated development for R*, version 1.2.5033. www.rstudio.com
Sarkar, Deepayan. 2018. *lattice: Trellis graphics for R*. R-package version 0.20-38. https://cran.r-project.org/web/packages/lattice/index.html
Schützler, Ole. 2015. *A sociophonetic approach to Scottish Standard English*. Amsterdam and Philadelphia: John Benjamins.

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

Schützler, Ole. 2020. Frequency changes and stylistic levelling of 'though' in diachronic and synchronic varieties of English – linguistic democratisation? *Language Sciences* 79.
Schützler, Ole, Ulrike Gut & Robert Fuchs. 2017. New perspectives on Scottish Standard English: Introducing the Scottish component of the International Corpus of English. In Joan C. Beal & Sylvie Hancil (eds.), *Perspectives on Northern Englishes*, 273–302. Berlin: Mouton de Gruyter [TiEL 96].
Schützler, Ole & Jenny Herzky. 2021. *Modal verbs of strong obligation in Scottish Standard English: Corpus data*. DataverseNO, V1, UNF:6:1qo2z6HvFxex9k9bSnXYLA==[fileUNF], https://doi.org/10.18710/GIP0PM
Smith, Jennifer. 2012. Scottish English and varieties of Scots. In Kortmann & Lunkenheimer (eds.), 21–9.
Smith, Nicholas. 2003. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In Facchinetti, Krug & Palmer (eds.), 241–66.
Sönning, Lukas. 2018. Drawing on principles of perception: The line plot. Paper given at ICAME39, Tampere.
Stan Development Team. 2019. *Stan modeling language user's guide and reference manual*, version 2.26. http://mc-stan.org
Sweetser, Eve E. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
Tagliamonte, Sali & Jennifer Smith. 2006. Layering, competition and a twist of fate: The deontic modality system in English dialects. *Diachronica* 23(2), 341–80.
Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 12(1), 49–100.
Tizón-Couto, David & David Lorenz. 2015. Everything matters, or what to do with all those variables… Paper given at the 5*th International Conference of the International Society for the Linguistics of English* (ISLE5), London.
Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

## Appendix

### *A. Corpus sizes: genres, numbers of texts and words*

Although ICE-GB was purged of extra-corpus material, the word count is rather too high for some text categories. Divergent figures for ICE-GB and ICE-SCO are due to the latter being still under construction. It also follows a somewhat different sampling approach, in that texts shorter than 2,000 words were accepted.

| | | | ICE-GB | | ICE-SCO | |
|---|---|---|---|---|---|---|
| Mode | Genre | abbrev. | $n_{texts}$ | $n_{words}$ | $n_{texts}$ | $n_{words}$ |
| SPOKEN | Broadcast discussions | bdis | 20 | 42,246 | 2 | 1,896 |
| | Broadcast interviews | bint | 10 | 21,647 | 9 | 5,699 |
| | Business transactions | btran | 10 | 19,558 | 1 | 454 |
| | Unscripted speeches | unsp | 30 | 64,918 | 15 | 12,569 |
| | Demonstrations | dem | 10 | 22,124 | 4 | 2,035 |
| | Legal presentations | leg | 10 | 22,098 | 14 | 9,084 |

(*Continued*)

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

(*continued*)

| Mode | Genre | abbrev. | ICE-GB $n_{texts}$ | ICE-GB $n_{words}$ | ICE-SCO $n_{texts}$ | ICE-SCO $n_{words}$ |
|---|---|---|---|---|---|---|
| | Broadcast news | bnew | 20 | 42,629 | 20 | 5,661 |
| | Broadcast talks | btal | 20 | 43,871 | 43 | 32,075 |
| | Non-broadcast talks | nbtal | 10 | 21,320 | 7 | 4,017 |
| WRITTEN | Social letters | sl | 15 | 31,090 | 38 | 3,661 |
| | Social sciences (academic) | ASoc | 10 | 23,135 | 3 | 6,900 |
| | Humanities (popular) | PHum | 10 | 21,147 | 11 | 8,419 |
| | Social sciences (popular) | PSoc | 10 | 30,751 | 11 | 5,702 |
| | Natural sciences (popular) | PNat | 10 | 30,937 | 19 | 9,275 |
| | Technology (popular) | PTec | 10 | 28,560 | 22 | 13,665 |
| | Press news reports | Rep | 20 | 44,223 | 79 | 40,265 |
| | Skills/hobbies | SkHo | 10 | 24,381 | 27 | 20,354 |
| | Press editorials | ed | 10 | 21,150 | 14 | 9,331 |
| | Novels & short stories | nov | 20 | 49,283 | 3 | 5,096 |
| TOTAL (spoken) | | | 140 | 300,411 | 115 | 73,490 |
| TOTAL (written) | | | 125 | 304,657 | 227 | 122,668 |
| **TOTAL** | | | **265** | **605,068** | **342** | **196,158** |

### *B. Model specification*

The model was set up as a multinomial regression model with four outcome categories, the reference category being the verb MUST. Priors were specified as documented in the full model syntax below. The model was run with $n = 4{,}750$ iterations in four chains, each with a warmup of $n = 1{,}000$ iterations. The resulting number of posterior samples was $n = 15{,}000$. Model diagnostics indicated the convergence of chains (R-hat = 1.00 for all parameters; see appendix C).

```
brm(verb ~ (subject + source + spoken) * scot
     + (subject + source | text)
    + (1 | genre),
family = "categorical", data = modals,
warmup = 1000, iter = 4750, cores = 3, chains = 4,
prior=c(
set_prior("normal(0, 3)", class = "Intercept"),
set_prior("normal(0, 2)", class = "b"),
set_prior("normal(0, 2)", class = "sd", dpar = "muhaveto"),
set_prior("normal(0, 2)", class = "sd", dpar = "muhavegotto"),
set_prior("normal(0, 2)", class = "sd", dpar = "muneedto")
),
control = list(adapt_delta = .96)
)
```

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

We tested the reported model against two models, one with random intercepts only for TEXT and GENRE, the other with the complete random-effects structure for TEXT but without random effects based on GENRE. Our model was evaluated as better than the less complex ones, based on the Leave-one-out Information Criterion (LOOIC).

To check for prior sensitivity, a model with relaxed priors was run, defining the prior of the intercept as 'normal(0, 6)' and the remaining four priors as 'normal(0, 3)'. This model was characterised by larger estimates for the standard deviations of several random coefficients. However, the visual inspection of predicted percentages did not reveal any differences that ran against our original conclusions, although uncertainties tended to increase. The model summary shown when applying the function print() to the model object in R is documented on the osf-repository (see section 4.4).

### C. *Regression table*

| | | | 95% CI | | $n_{eff}$ | | |
|---|---|---|---|---|---|---|---|
| Outcome | Coefficient | Mean | Lower | Upper | bulk | tail | R−hat |
| HAVE TO | INTERCEPT | −0.00 | −0.68 | 0.63 | 9,291 | 9,883 | 1.00 |
| | SUBJECT (2nd) | 1.14 | −0.03 | 2.43 | 11,052 | 10,451 | 1.00 |
| | SUBJECT (3rd) | −0.29 | −0.93 | 0.33 | 10,746 | 11,327 | 1.00 |
| | SOURCE (subj.) | −0.92 | −1.30 | −0.56 | 9,577 | 10,262 | 1.00 |
| | SPOKEN | 0.70 | 0.20 | 1.22 | 7,798 | 9,138 | 1.00 |
| | SCOT | 1.02 | 0.11 | 1.92 | 10,230 | 10,488 | 1.00 |
| | SUBJECT (2nd):SCOT | 0.62 | −1.76 | 3.03 | 15,517 | 12,166 | 1.00 |
| | SUBJECT (3rd):SCOT | −1.18 | −2.33 | −0.07 | 9,795 | 9,892 | 1.00 |
| | SOURCE (subj.):SCOT | −0.25 | −0.95 | 0.44 | 10,974 | 11,502 | 1.00 |
| | SPOKEN (subj.):SCOT | −0.03 | −0.64 | 0.58 | 12,070 | 11,003 | 1.00 |
| NEED TO | INTERCEPT | −1.32 | −2.32 | −0.44 | 5,162 | 7,309 | 1.00 |
| | SUBJECT (2nd) | 0.48 | −1.46 | 1.96 | 5,770 | 8,068 | 1.00 |
| | SUBJECT (3rd) | −0.63 | −1.64 | 0.35 | 5,363 | 8,546 | 1.00 |
| | SOURCE (subj.) | −0.18 | −0.67 | 0.32 | 7,586 | 8,910 | 1.00 |
| | SPOKEN | 0.10 | −0.57 | 0.75 | 7,758 | 9,568 | 1.00 |
| | SCOT | 2.43 | 1.43 | 3.51 | 7,596 | 8,925 | 1.00 |
| | SUBJECT (2nd):SCOT | 0.33 | −2.01 | 2.81 | 15,409 | 11,605 | 1.00 |
| | SUBJECT (3rd):SCOT | −0.97 | −2.20 | 0.27 | 7,430 | 9,688 | 1.00 |
| | SOURCE (subj.):SCOT | 0.09 | −0.65 | 0.83 | 10,939 | 10,530 | 1.00 |
| | SPOKEN (subj.):SCOT | 0.32 | −0.37 | 0.99 | 11,456 | 11,016 | 1.00 |
| (HAVE) GOT TO | INTERCEPT | −4.12 | −6.42 | −2.20 | 7,675 | 9,112 | 1.00 |
| | SUBJECT (2nd) | 0.69 | −1.40 | 2.46 | 9,837 | 10,104 | 1.00 |
| | SUBJECT (3rd) | −0.79 | −2.21 | 0.47 | 10,333 | 9,409 | 1.00 |
| | SOURCE (subj.) | −1.05 | −1.84 | −0.29 | 11,238 | 10,570 | 1.00 |
| | SPOKEN | 1.18 | −0.40 | 2.84 | 7,087 | 8,386 | 1.00 |
| | SCOT | −0.81 | −2.80 | 0.98 | 14,803 | 11,671 | 1.00 |
| | SUBJECT (2nd):SCOT | 0.30 | −2.78 | 3.31 | 16,593 | 11,862 | 1.00 |

(*Continued*)

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press

(*continued*)

| Outcome | Coefficient | Mean | 95% CI Lower | 95% CI Upper | $n_{eff}$ bulk | $n_{eff}$ tail | R−hat |
|---|---|---|---|---|---|---|---|
| | SUBJECT (3rd):SCOT | −1.07 | −3.53 | 1.28 | 15,063 | 12,116 | 1.00 |
| | SOURCE (subj.):SCOT | 0.49 | −1.08 | 2.10 | 14,056 | 12,083 | 1.00 |
| | SPOKEN (subj.):SCOT | 0.39 | −1.18 | 1.97 | 14,374 | 12,052 | 1.00 |
| TEXT: standard deviations of varying intercepts and slopes ($n = 332$) | | | | | | | |
| HAVE TO | INTERCEPT | 0.89 | 0.13 | 1.55 | 1,214 | 1,154 | 1.00 |
| | SUBJECT (2nd) | 2.56 | 0.40 | 4.99 | 1,754 | 1,741 | 1.00 |
| | SUBJECT (3rd) | 0.63 | 0.03 | 1.54 | 1,132 | 3,500 | 1.00 |
| | SOURCE (subj.) | 0.81 | 0.09 | 1.47 | 1,613 | 2,271 | 1.00 |
| NEED TO | INTERCEPT | 1.10 | 0.14 | 2.09 | 1,821 | 2,146 | 1.00 |
| | SUBJECT (2nd) | 2.41 | 0.25 | 4.90 | 2,222 | 3,553 | 1.00 |
| | SUBJECT (3rd) | 1.30 | 0.11 | 2.72 | 1,210 | 2,968 | 1.00 |
| | SOURCE (subj.) | 0.60 | 0.03 | 1.37 | 2,035 | 4,704 | 1.00 |
| (HAVE) GOT TO | INTERCEPT | 1.49 | 0.23 | 2.75 | 2,252 | 2,665 | 1.00 |
| | SUBJECT (2nd) | 2.67 | 0.31 | 5.19 | 3,165 | 2,996 | 1.00 |
| | SUBJECT (3rd) | 0.92 | 0.04 | 2.45 | 3,105 | 6,372 | 1.00 |
| | SOURCE (subj.) | 0.80 | 0.04 | 1.90 | 3,326 | 6,221 | 1.00 |
| GENRE: standard deviations of varying intercepts ($n = 19$) | | | | | | | |
| HAVE TO | INTERCEPT | 0.86 | 0.47 | 1.40 | 6,344 | 9,255 | 1.00 |
| NEED TO | INTERCEPT | 1.05 | 0.49 | 1.80 | 5,928 | 7,269 | 1.00 |
| (HAVE) GOT TO | INTERCEPT | 2.94 | 1.60 | 4.75 | 8,006 | 9,071 | 1.00 |

https://doi.org/10.1017/S1360674321000071 Published online by Cambridge University Press