The morpho-syntax of Standard Scottish English: Questionnaire-based insights

Ole Schützler (Leipzig University)
Lukas Sönning (University of Bamberg)
Fabian Vetter (University of Bamberg)
Manfred Krug (University of Bamberg)

Corresponding author: ole.schuetzler@uni-leipzig.de

Abstract. With empirical research on Scottish Standard English showing a bias towards phonology, the aim of the current study is to contribute to the morpho-syntactic documentation of this variety. Relying on Standard Southern British English for comparison, we concentrate on four eWAVE features on which the varieties have been noted to diverge: (i) *youse* as a second person plural pronoun, (ii) extended uses of the progressive, (iii) epistemic *mustn't*, and (iv) quotative *like*. We use questionnaire data, where respondents indicate how many speakers in their home country use a particular feature. Ratings are elicited from 43 English and 61 Scottish participants (mostly university students) for two usage contexts, (informal) speech and (semi-formal) writing. Our findings corroborate expert ratings in eWAVE and show that the reported currency of all features is higher in Scottish Standard English, albeit to varying degrees. Our study draws attention to the complementary potential of questionnaire data in World Englishes research.

Keywords: Standard Scottish English; Standard Southern British English; morpho-syntax; questionnaire data; eWAVE; *yous(e)*; progressive; epistemic *mustn't*; quotative *like*

1. Introduction

In comparison to other standard varieties of English, Standard Scottish English (SSE)[1] has experienced a certain empirical neglect in the literature and therefore remains a relatively underdocumented variety, as argued by Schützler (2024), who outlines some of the underlying historical, political and linguistic reasons. In existing quantitative work on SSE there tends to be a bias towards phonological phenomena. The particularly pronounced gap at the level of morpho-syntax may in part be due to the commonly held belief that standard varieties vary little at these levels of linguistic description (McArthur 1987). Uniquely to the Scottish context, however, this may be aggravated by a tendency to interpret interesting phenomena as features of Scots, not English (Schützler, Gut & Fuchs 2017), and by the fact that SSE is at the same time a variety of British English, so that standard features (and the standard variety as a whole) may implicitly be regarded as shared.

The aim of the current study is to enrich the linguistic description of SSE and to contribute towards filling the morpho-syntactic void in the empirical literature. Relying on Standard Southern British English (SSBE) as a point of reference, we concentrate on four features drawn from eWAVE (Kortmann et al. 2020), regarding which SSE has been impressionistically noted to diverge from SSBE: (i) *youse* as a second person plural pronoun, (ii) extended use of the progressive with stative verbs, (iii) epistemic *mustn't*, and (iv) quotative *like*. The current study relies on questionnaire data, which are drawn from the *Bamberg Survey of Language Variation and Change* (BSLVC; Krug and Sell

[1] We intentionally depart from the currently more usual label *Scottish Standard English* to emphasize that we are looking at the standard pole of an internally fully differentiated Scottish variety of English, not a Scottified variant of (World) Standard English. This has no practical implications for our research.

2013). In this pen-and-paper survey, respondents are asked to indicate how many speakers in their home country use a particular feature (*no-one*, *few*, *some*, *many*, *most*, *everyone*). This allows us to draw attention to advantages of questionnaires vis-à-vis corpora for the study of structures that are infrequent, difficult to detect in textual data, or which require a nontrivial choice of baseline (i.e. choice contexts) to allow for meaningful comparisons between varieties (see Wallis and Mehl 2022).

The paper is structured as follows. In Section 2, we briefly review the current state of research on SSE (morpho-syntax) to motivate the present study. Section 3 then lays out the morpho-syntactic features we focus on, and reviews the existing literature regarding their prevalence in British varieties, with a focus on Scottish English. Our methods and data are presented in Section 4, followed in Section 5 by a presentation of our findings. Section 6 offers a summary and discussion of our empirical results and in Section 7 we recapitulate and weigh the advantages and disadvantages of questionnaire data and their role in complementing research on varieties of English.

2. Standard Scottish English: An underdocumented variety

This section describes two ways in which the documentation of SSE may be considered inadequate. As such, SSE has not received the same attention as other standard varieties of English, leading to an underrepresentation in the literature on World Englishes. Further, the state of research on the characteristics of SSE is imbalanced, as there is a notable bias towards phonological structures. We will deal with these aspects in turn.

Corbett, McClure and Stuart-Smith (2003: 4) consider the inadequate documentation of SSE as resulting from its intermediate position between two salient varieties, Broad Scots and SSBE. Schützler (2024) also gives socio-political reasons for the state of affairs, including what he calls 'the narrative of loss' (i.e. a focus on the erosion of salient features of Scots in the evolution of SSE, rather than the maintenance or development of features), the 'Scots Bias' (i.e. a pronounced engagement in the literature with this historically and linguistically more clearly defined variety), and the fact that standard varieties are traditionally associated with independent nation states.

As for the attention directed at different levels of language description, Schützler (2024) observes a pronounced bias towards phonological features. This empirical imbalance may be partly attributable to the general observation that standard varieties tend to be relatively closely aligned in terms of morphology and syntax and that the push towards uniformity is particularly strong in writing (McArthur 1987). It has thus been repeatedly stated for SSE that it resembles SSBE quite closely at these levels of linguistic description (Wells 1982: 394; Giegerich 1992: 45-46; see also Svartvik and Leech 2016: 147). It deserves to be stressed, however, that this bias does not seem to exist for other standard varieties, at least not to the same extent. This is despite the fact that they, too, are structurally quite similar to each other. In the case of SSE, Scotland's geographical proximity to England and the lack of political independence are presumably factors that lead to an underestimation or even oversight of characteristic features.

The linguistic literature started to engage with Standard Scottish English in the late 1970s. An early treatment is offered by Aitken (1979), who observes that SSE appeared to be more distinct in terms of grammar than previously assumed. McClure (1994: 85) was perhaps the first to comment on the empirical gap that exists for non-phonological features, noting that 'here it is more difficult to obtain precisely quantifiable data'. In the era of corpora, this difficulty may in large part be attributed to the absence of textual databases. As discussed by Schützler, Gut and Fuchs (2017: 279), publicly available corpora such as the Scottish Corpus of Texts & Speech (SCOTS; Douglas 2003; Anderson and Corbett 2008), and the Corpus of Modern Scottish Writing (CMSW) provide a valuable basis for the

isolated description of linguistic features. However, due to the lack of comparability with corpora on other varieties including SSBE, it is more difficult to situate SSE in relation to other standards of English. The recent initiative to add a Scottish component to the family of ICE corpora (International Corpus of English, Greenbaum and Nelson 1996) aims to broaden the empirical opportunities for research on SSE morpho-syntax. For an ICE-based study comparing the use of modals of strong obligation in SSE and SBSE, see Schützler and Herzky (2021).

3. Morpho-syntactic features: Background

This section offers background on the morpho-syntactic structures studied, dealing with each feature in turn. We review the literature and take a look at data from *The Electronic World Atlas of Varieties of English* (eWAVE; Kortmann et al. 2020) to map their currency in certain British English dialects. Section 3.5 provides a summary. The examples we use for illustration are the questionnaire items that were rated by our participants (see Section 4.1).

3.1. Yous(e)

The dialectal form *yous(e)* (variously spelled *yous* or *youse*[2]) establishes (or *re*-establishes) number contrast for second-person pronouns and therefore fills a paradigmatic gap that exists in the modern (standard) English pronominal system. Thus, in (1) the speaker makes it clear that they suggest two or more people join in:

(1) Why don't **youse** come along to the restaurant?

The use of *yous(e)* as a second person plural pronoun appears to have its origin in Irish English (likely reflecting substrate effects of Gaelic), from where it has spread through migration flows, mainly during the 19th century (Beal 1997: 344-346). It is therefore found in regions that experienced significant immigration of Irish people. For a general overview of its currency in British dialects, the first column in Table 1 collects the ratings given in eWAVE. The feature is rated as pervasive or obligatory in Ireland and Northern England (A rating), which is consistent with its assumed evolution and diffusion. For Scottish English, *yous(e)* receives a B rating, classifying it as "neither pervasive nor extremely rare".

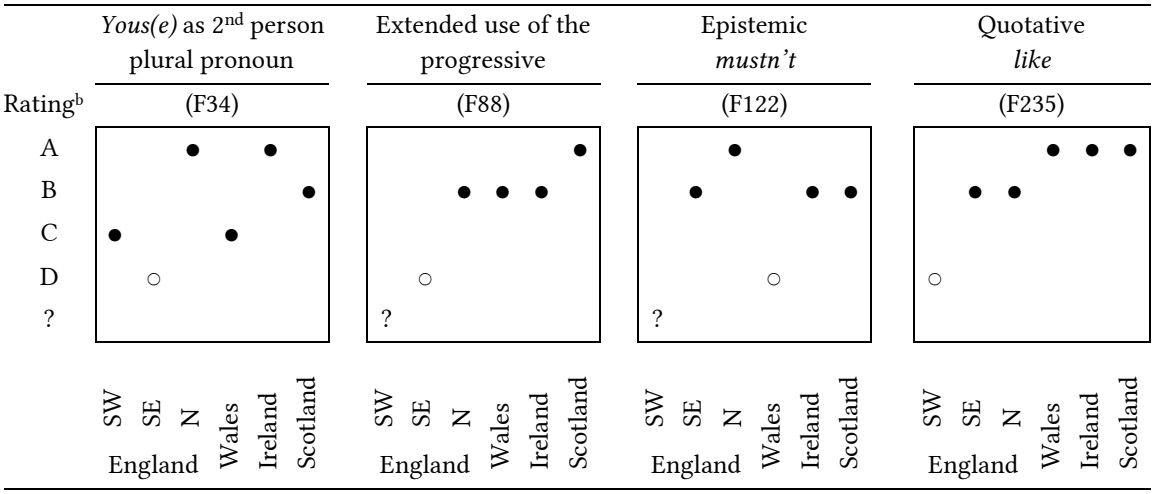
A survey by MacKenzie et al. (2022) corroborates the eWAVE records, showing that *yous(e)* is used at a high rate in the northeast of England and in Scotland. The web-based questionnaire, which mainly reached younger cohorts (median age 22 years, 39% self-declared students) asked participants "How would you address a group of two or more people?", the response options being *you*, *you guys*, *yous(e)*, and *you lot*. Among the 263 Scottish informants who responded to that question, 33% selected *yous(e)*. This relatively high currency is consistent with earlier accounts of this feature in Scottish English.

As for the regional distribution of this feature within Scotland, Smith (2012: 22) observes that it "is used extensively in the Central Belt of Scotland: Edinburgh, Glasgow and surrounding areas", and Crawford (2017) likewise considers it a feature of urban Central Scots dialects. Maguire (2012: 70-71) presents data from the Linguistic Survey of Scotland, which show "a clear south-westerly bias in its distribution" in traditional Scots varieties. Corbett and Stuart-Smith (2012) also report that *yous(e)* is relatively common in mainland Scotland, and they found 37 instances in the 800,000-word-long

[2] Other spellings are *yis* or *yus*, and Pearce (2021: 199) lists further variants found in the RTG corpus, based on an online forum associated with a football club in North East England: *yous*, *youse*, *yuz*, *yez*, *yees*, *yeez*, *yerz*, *yes*, *yiz*, *yays*.

spoken part of the SCOTS corpus (Scottish Corpus of Texts and Speech). As for its social distribution, Miller (1993: 108) observes that *yous(e)* is "assiduously avoided by educated speakers even in informal situations". In a later publication, Miller (2008: 301) repeats this statement in a shorter, weakened form ("avoided by educated speakers"); this aligns with the first author's impression that this form is now regularly heard in everyday spoken interaction, even if it may still not be considered a standard feature. Beal (1993: 205), on the other hand, notes that some younger speakers also use *yous(e)* when addressing one person, indicating that the form may be perceived as the local form of the second person pronoun.

Table 1. The features[a] under study: Currency ratings[b] from eWAVE[c] (Kortmann et al. 2020).



Note. [a] The eWAVE features are labeled as follows (italics added): (F34) "forms of phrases for the second person plural pronoun other than *you*"; (F88) "wider range of uses of progressive *be* + V-*ing*: extension to stative verbs"; (F122) "epistemic *mustn't*"; (F235) "*like* as a quotative particle"

[b] Key to the ratings: (A) feature is pervasive or obligatory; (B) feature is neither pervasive nor extremely rare; (C) feature exists, but is extremely rare; (D) attested absence; (?) no information available

[c] References for the currency ratings: Scottish English (Smith 2020), Irish English (Filppula 2020), Welsh English (Penhallurick 2020), dialects in the north (Trousdale 2020), southwest (Wagner 2020) and southeast (Anderwald 2020) of England.

3.2. Extension of the progressive to stative verbs

The prototypical use of the BE + V-*ing* construction as a marker of progressive aspect is to signal that the action denoted by the verb is ongoing (or in progress). It is therefore chiefly associated with dynamic verbs, which are capable of having a duration. Since the 17th century, however, constraints on the use of the progressive have loosened and its scope has since widened to other situation types expressed by stative verbs (Kranich 2010). Non-dynamic verbs are then coerced into the imperfective aspect, which means that they are given a dynamic interpretation (Michaelis 2004). As a result, the progressive has acquired a number of additional, non-prototypical meanings. For instance, in (2) it can be interpreted as intensifying, while in (3) it can be understood as a politeness strategy:

(2) I'm really **liking** this film.
(3) What are you **wanting**?

A number of corpus-based studies have documented the extended use of the progressive in British English (e.g. Leech et al. 2009; Levin 2013), but there are indications that its spread may have halted

(Rautionaho and Fuchs 2021). Corbett and Stuart-Smith (2012: 90) note that "Scottish English has long allowed a greater range of verbs in the progressive form than was conventionally found in other varieties of standard English" and the extended use of progressive aspect is a commonly identified feature of Scots and Scottish English (e.g. Aitken 1979: 106; Beal 1997: 372-373; Macafee and Ó Baoill 1997: 269-273). Miller (2008: 307) notes that in Scots, stative verbs other than *know* regularly appear in the progressive. Smith (2012: 23) identifies the use of the progressive with stative verbs as a covert feature that is widespread in Scottish varieties and also used by speakers of SSE. This is consistent with observations made by Miller (2008: 307), who cites examples of stative progressives (*liking, wanting, understanding, intending*) from Scottish educated written usage (undergraduate essays). The expert judgments in eWAVE also suggest that this feature has made significant inroads into Scottish English. As column 2 in Table 1 shows, it is the only dialect (in our set of six) that receives an A rating for this feature, indicating that it is pervasive or obligatory.

3.3. Epistemic *mustn't*

In standard English, the modal verb *must* has two senses: (i) a deontic sense, which expresses obligation and typically involves some form of human control over events, and (ii) an epistemic sense, which expresses necessity (or "conclusion"), i.e. human judgement about the truth-value of some proposition. Miller (1993: 117; 2008: 305) states that in Scots *must* only has the epistemic "conclusion" meaning; *mustn't* therefore expresses 'I conclude that not'. This use of epistemic *mustn't* is illustrated in (4), where the speaker is left to conclude that something is false:

(4) It's so unlikely, it **mustn't** be true.

Epistemic *mustn't* is also a feature of northern English dialects (Kortmann 2008; Beal 2010: 38), and Kortmann (2008: 491) lists it as a feature distinguishing northern and southern English dialects. As column three in Table 1 indicates, the accounts given in eWAVE largely support the classification of epistemic *mustn't* as a feature of dialects in the north of England. In Scotland, it is labeled as "neither pervasive nor extremely rare" (B rating).

3.4. Quotative *like*

The quotative use of *like*, i.e. to introduce thought or reported speech, has first been noted in the early 1980s (Butters 1980), and has since seen a "meteoric rise" (Gardner et al. 2021: 282) and rapid spread across varieties of English (Buchstaller and D'Arcy 2009). In the course of this development, it has lost its social stigma ("Valley Girl" reputation; see Blyth et al. 1990) and expanded its semantic scope beyond the original dramatized personal narrative (Ferrara and Bell 1995). As a result, it competes with other quotative devices (e.g. *say, go, think*). In (5), for instance, it introduces direct speech:

(5) And then she **was like** "What do you mean?"

Its common occurrence in British dialects is well-attested, as illustrated in Table 2, which is based on the literature review offered by Britain (2020: 40). The currency reports listed there express the share of *be like* among all quotative devices, and they are arranged by latitude (northern-most sites appear at the top) and decade. Since real- and apparent-time studies have recorded considerable increases in usage (Durham et al. 2011; Gardner et al. 2021), the figures cited in Table 2 concentrate on younger speaker groups. There is some indication in Table 2 that the usage rate of quotative *like* is slightly higher in northern varieties. Column four of Table 1 shows that according to eWAVE, quotative *like* appears to be characteristic of Scottish English (as well as Welsh and Irish Englishes), where it is

reported to enjoy wider usage (A rating) compared to dialects in the north and south-east of England (B rating).

Table 2. Usage rate of quotative *like* across British dialects (largely based on Britain 2020).

			Usage rate	
Site	Demographics	Reference	1990s	2000s
Edinburgh	Teenagers	Meyerhoff & Schleef 2014		47%
York	Undergraduate students	Tagliamonte & Hudson 1999	18%	
York	Undergraduate students	Durham et al. 2011		68%
Newcastle/Derby	Young speakers (14-26)	Buchstaller 2006	4.5%	
Leeds	Students (school)	Richards 2008		59%
London	Teenagers	Stenström et al. 2002	<1%	
Inner London	Young speakers	Cheshire et al. 2011		24%
Outer London	Young speakers	Cheshire et al. 2011		21%

Meyerhoff and Schleef (2014) studied the use of *like* as a quotative verb in 21 Edinburgh-born teenagers, based on recordings of free conversation. Overall, *be like* was the most frequent reporting expression (47% of instances, compared to 21% *say*, 17% zero, and 12% *go*), indicating that it has gained considerable ground in this population of speakers. Macaulay (2005: 82) reports on quotative *like* in Glasgow adolescent conversations and observed social class differences: The usage rate (per one million words) was higher in middle-class (vs. working class) adolescents, and girls showed higher rates than boys in both subgroups.

3.5. Summary

According to our literature survey, the four morpho-syntactic features under study are associated with dialects in the north of England and/or Scottish English or Scots. We would therefore expect them to be more prevalent in SSE compared to SSBE. Even if some of the features – particularly plural *yous(e)* and quotative *like* – traditionally carried some stigma that would have disqualified them as features of the standard dialect, there is evidence that this stigma has worn off to a considerable extent. As the four features are innovative (or even non-standard), we would further expect them to enjoy relatively more widespread use in informal speech than in more formal styles of writing.

4. Method and data

We now turn to the methods and data we use to compare SSE and SSBE. Section 4.1 provides background on the BSLVC, from which the data are drawn, and Section 4.2 describes our sample of informants. In Section 4.2, we outline our statistical approach.

4.1. Questionnaire design and administration

The current dataset is drawn from the *Bamberg Survey of Language Variation and Change* (BSLVC), a large-scale survey on the use of lexical and grammatical structures in different varieties of English. It was initiated in 2008 with a focus on Romance-English language contact (Malta, Gibraltar, Channel Islands), and its scope has since widened to other sites, including England, Wales, Scotland, Australia, Puerto Rico and the US, and to regions where English is spoken as a foreign language (i.e.

EFL varieties, including Germany, Slovenia and Sweden). As outlined in Krug and Sell (2013: 80), the questionnaire is made up of four parts. Part (I) consists of an informant sheet asking for personal information such as age, gender, nationality, ethnic self-identification, languages used at home while growing up, regional background, and an education and occupation profile. Part (II) elicits usage ratings for 138 spoken sentences (target register: informal conversation among friends), which contain a broad range of morpho-syntactic and pragmatic features. Part (III) investigates lexical preferences for 68 British-American (near-)synonyms (e.g. *pavement-sidewalk, diapers-nappies*). Part (IV) elicits usage ratings for 207 written sentences (target register: email to a former school teacher), again focusing on morpho-syntactic features and including the 138 sentences of part (II).

The main target group of the grammar part of the BSLVC have thus far been university and secondary school students, and the questionnaires from which the current dataset is drawn were typically administered with relatively large groups in university classrooms. This approach was partly motivated by two practical considerations: (1) A quiet environment is required for the auditory presentation of sentences in part II (see below), and (2) collecting data from multiple participants using a single run reduces the considerable expenses involved in the administration of the 80-minutes-long full questionnaire. Mainly approaching a subpopulation of students seemed reasonable because in many second-language varieties it is only the more highly educated strata of society that actually use English to a considerable extent. This also brings the BSLVC into closer alignment with L2 components of the International Corpus of English (ICE; Greenbaum & Nelson 1996), which focus on educated English rather than on all levels of competence.

The current study draws on usage ratings collected in parts (II) and (IV). In part (II), 138 sentences are presented twice to the participants (identical sound files produced by a native speaker of the respective variety). Participants are asked to rate whether each sentence could be said in their home country in an informal conversation among friends. Responses are given on a sheet of paper, by ticking one of the following options: *everyone, most, many, some, few, no-one*. In part (IV), 207 sentences are presented in writing, printed on the questionnaire form. Using the same 6-point scale, respondents are again asked to indicate how prevalent a feature is in their home country or region. To elicit ratings for a semi-formal register, participants were asked to consider the usage context of an email to a former teacher.

4.2. Participants

In total, 104 individuals participated in the study. The 61 Scottish participants (24 female, 37 male) are between 18 and 33 years of age (mean age: 21). Figure 1 shows their regional background, i.e. the city or part of the country where they spent most of their time while growing up and where they went to secondary school. Each filled circle represents a respondent, with the exception for the two major cities, Edinburgh (n = 12) and Glasgow (n = 15), where a larger symbol is used to denote multiple speakers. Figure 1 shows that most of our Scottish participants are from the Central Belt, i.e. Edinburgh, Glasgow and surrounding areas. The 43 English participants (21 female, 22 male) are between 18 and 24 years old (mean age: 21), and they are spread quite widely across England. Around 20 participants are from southern England, and another cluster of participants may be found where the Midlands border on Yorkshire and Northern England. The data used in the present study are available from TROLLing (Krug et al. 2025) and the R scripts for reproducing the analyses are deposited to the OSF (https://osf.io/hykq5/).

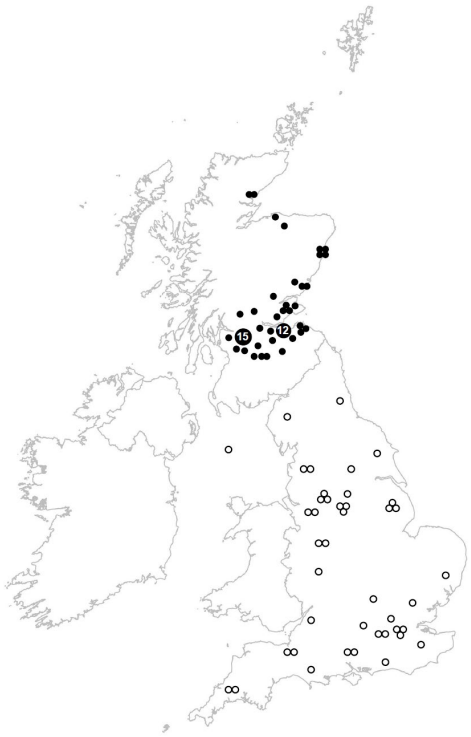


Figure 1. Regional background of participants.

4.3. Statistical analysis

The usage rates reported by our participants are collected using an ordinal response scale. For the statistical analysis of our data, we take advantage of the fact that the quantifiers used to label the response options can be interpreted as relative frequencies: *No-one* and *everybody* mark the endpoints of the percentage scale (0% and 100%), and the intermediate expressions (*few*, *some*, *many*, *most*) likewise have a proportional meaning. A number of studies have used experimental techniques to pin down the typical percentage denoted by these verbal quantifiers (e.g. Newstead et al. 1987). As discussed in more detail in Sönning (2024), a survey of the literature suggests that the typical quantificational meaning of our response labels are as follows: *few* (11%), *some* (27%), *many* (67%), *most* (81%).

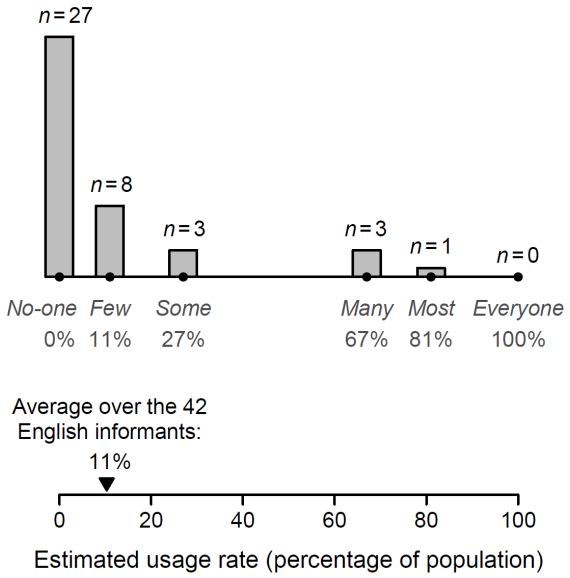


Figure 2. Illustration of the custom scoring system and its use for the analysis and interpretation of the data: semi-formal *yous*(*e*), English participants.

For the statistical analysis of our data, we will replace the response categories with these proportional values. This is illustrated in Figure 2, which uses the usage reports by our English participants for *yous(e)* in semi-formal writing. The bars show the distribution of responses across the six categories; for instance, 8 individuals reported that "few" people in their home country would use this sentence in an email to a former teacher. The horizontal spacing of the bars shows that the gap between 'some' and 'many' is stretched, in line with our knowledge about how speakers interpret these expressions. If we translate responses into their proportional analogues, we can average over the 42 scores, which yields a value of 11%. We will interpret this average as an estimate of the usage rate of *yous(e)* in the population: According to the participants of our study, roughly 1 out of 10 speakers of SSBE would use *yous(e)* in semi-formal writing. These percentages provide a useful frame of reference for the summary and interpretation of the BSLVC data. We will read them as rough estimates of the overall prevalence of a feature in the varieties and contexts under investigation.

The use of a numeric-conversion approach to the analysis of ordinal data is controversial (see, e.g., Knapp 1990). As discussed in more depth in Sönning (2024), measurement-theoretic and statistical issues must be given due consideration when deciding on an analysis strategy for ordinal data. As laid out in more detail there, from a measurement-theoretic perspective the goals of our analysis grant us some leeway in our choice of analysis strategy. Since we do not intend to advance causal interpretations of the patterns in our data, but are merely interested in describing and comparing them, the much-cited prescriptions (e.g. Stevens 1946) do not apply to the ensuing analyses. Further, since the numeric scores (i.e. percentages) we assign to the response labels are grounded in experimental research, the distances we assume to hold between these verbal anchors are not arbitrarily chosen but motivated empirically.

As for the statistical reservations that apply to the numeric-conversion approach to ordinal data, we will sidestep issues resulting from the boundedness of the response scale by applying a form of logistic regression. Note that this is only possible because our ordinal response variable has a proportional interpretation, which can be expressed on the [0, 100] scale. More specifically, we use fractional regression (e.g. Papke and Wooldridge 1996; see also Kieschnick and McCullogh 2003;

Crowell and Fossett 2017) to model the proportional data on an unconstrained logit scale. This can be done in R (R Core Team 2025) using the *glm()* function with the argument 'family' set to 'quasibinomial' (see Clark 2019). We further used the package *marginaleffects* (Arel-Bundock 2023) for constructing model-based predictions and comparisons on the percentage (rather than logit) scale, and relied on the package *lattice* (Sarkar 2008) for data visualization.

Identical models[3] were used for each feature, with main effects for both binary predictors – Variety and Mode – as well as an interaction between them. We decided to retain the interaction in each model for two reasons: First, fractional regression examines statistical interactions on the logit scale, which is not the scale on which we seek to interpret our data. Rather, we are interested in proportional summaries, and whether an interaction is apparent on *this* scale. This means that the question addressed by the interaction coefficient in our regression model is of no direct interest. For this reason, we decided to specify our model in a way that returns condition-specific estimates (i.e. Scottish writing, Scottish speech, etc.), and we examine the question of statistical interaction using model-based predictions and comparisons on the percentage scale. To this end, we obtain estimated differences between the varieties (i) overall, i.e. averaging over mode, and (ii) separately for informal speech and for semi-formal writing. In addition, we will estimate what we will refer to as stylistic clines, which is the difference (expressed in percentage points) between (informal) speech and (semi-formal) writing. We therefore further estimate (iii) the stylistic cline for each variety, and finally (iv) the difference between the stylistic clines. It is this last quantity that addresses the question of a statistical interaction, but on the scale that forms the basis of our interpretation. 95% confidence intervals then allow us to appreciate the statistical uncertainty surrounding these estimated differences.

5. Results

We now turn to the results of our analyses, treating the four features in turn. Each section starts with a descriptive summary of the data, in the form of tables and graphs, and then presents the results from the regression analysis, also both in tabular and visual form. Tables with regression coefficients are deferred to the online materials (https://osf.io/5r8km/).

5.1. *Yous(e)* as a second person plural pronoun

We start with an overview of the frequencies of response categories in our data. Table 3 shows the distribution of the reported usage rates by Variety and Mode. For instance, upon hearing the sentence "Why don't youse come along to the restaurant?", 1 of the 60 Scottish participants who responded to this item (i.e. 2%) ticked the option "no-one", meaning that, in their impression, no one in their home country or region would say this sentence in an informal conversation among friends. For the English respondents, $n = 10$ selected this option, which corresponds to a greater proportional share of the sample (namely 10/43, or 23%).

[3] The only exception is the model for the extended use of the progressive, which includes Item as an additional factor (main effect only), since two questionnaire items elicited data on stative progressives.

Table 3. Distribution of responses by Variety and Mode.

	No-one		*Few*		*Some*		*Many*		*Most*		*Everyone*		Total	
Variety, Mode	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Yous(e)														
Scottish														
Speech	1	**2**	4	**7**	6	**10**	18	**30**	17	**28**	14	**23**	60	**100**
Writing	13	**21**	12	**20**	10	**16**	11	**18**	10	**16**	5	**8**	61	**100**
English														
Speech	10	**23**	10	**23**	13	**30**	6	**14**	3	**7**	1	**2**	43	**100**
Writing	27	**64**	8	**19**	3	**7**	3	**7**	1	**2**	0	**0**	42	**100**
Extended use of the progressive														
Scottish														
Speech	0	**0**	3	**2**	4	**3**	18	**15**	41	**34**	56	**46**	122	**100**
Writing	3	**3**	7	**6**	10	**8**	17	**14**	32	**27**	51	**43**	120	**100**
English														
Speech	8	**9**	7	**8**	12	**14**	18	**21**	22	**26**	18	**21**	85	**100**
Writing	9	**11**	16	**19**	16	**19**	16	**19**	18	**21**	10	**12**	85	**100**
Epistemic *mustn't*														
Scottish														
Speech	1	**2**	8	**13**	9	**15**	13	**21**	12	**20**	18	**30**	61	**100**
Writing	3	**5**	4	**7**	10	**17**	12	**20**	13	**22**	18	**30**	60	**100**
English														
Speech	3	**7**	4	**9**	11	**26**	4	**9**	14	**35**	6	**14**	42	**100**
Writing	5	**12**	7	**16**	6	**16**	10	**23**	9	**21**	5	**12**	42	**100**
Quotative *like*														
Scottish														
Speech	0	**0**	1	**2**	1	**2**	9	**15**	29	**48**	21	**34**	61	**100**
Writing	3	**5**	10	**16**	8	**13**	12	**20**	15	**25**	13	**21**	61	**100**
English														
Speech	0	**0**	0	**0**	6	**15**	21	**51**	11	**27**	3	**7**	41	**100**
Writing	5	**12**	11	**26**	13	**30**	5	**12**	6	**14**	3	**7**	43	**100**

The category percentages are shown graphically in Figure 3, using a diverging bar chart (Heiberger and Robbins 2014). Shades of red suggest low prevalence and shades of blue high prevalence of the feature in question. We observe a greater share of blue categories for the Scottish subgroup, and within each variety there is a stylistic cline, with a lower prevalence of features in moderately formal writing.

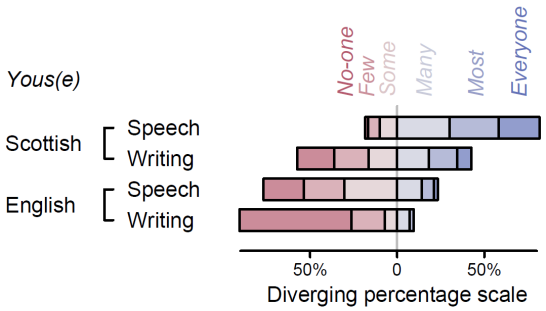


Figure 3. Distribution of ratings for *yous(e)*: Diverging bar chart.

Table 4. Model-based predictions and comparisons with (standard errors) and [95% confidence intervals].

Estimate	Average	Informal Speech	Semi-formal Writing	Stylistic cline (speech – writing)
Yous(e)				
Scottish participants	55.4 (2.9) [49.8, 61.1]	70.4 (3.9) [62.7, 78.1]	40.5 (4.2) [32.3, 48.7]	29.9 (5.8) [18.6, 41.2]
English participants	19.5 (2.8) [14.0, 25.0]	28.2 (4.6) [19.2, 37.2]	10.8 (3.2) [4.5, 17.0]	17.4 (5.6) [6.5, 28.3]
Difference (Scot. – Eng.)	35.9 (4.0) [28.1, 43.8]	42.2 (6.0) [30.4, 54.0]	29.7 (5.3) [19.4, 40.0]	12.5 (8.0) [−3.2, 28.2]
Extended use of the progressive				
Scottish participants	81.0 (1.6) [77.8, 84.1]	84.8 (2.1) [80.8, 88.9]	77.1 (2.4) [72.4, 81.9]	7.7 (3.2) [1.4, 14.0]
English participants	55.5 (2.4) [50.8, 60.2]	61.7 (3.3) [55.1, 68.2]	49.2 (3.4) [42.5, 56.0]	12.5 (4.8) [3.1, 21.8]
Difference (Scot. – Eng.)	25.5 (2.9) [19.9, 31.2]	23.1 (3.9) [15.4, 30.9]	27.9 (4.2) [19.6, 36.2]	−4.8 (5.8) [−16.1, 6.5]
Epistemic *mustn't*				
Scottish participants	66.1 (3.0) [60.2, 72.0]	65.5 (4.3) [57.2, 73.9]	66.6 (4.3) [58.2, 75.0]	−1.1 (6.0) [−12.9, 10.8]
English participants	53.9 (3.8) [46.4, 61.4]	56.5 (5.4) [45.9, 67.0]	51.3 (5.4) [40.7, 62.0]	5.1 (7.6) [−9.8, 20.1]
Difference (Scot. – Eng.)	12.2 (4.9) [2.6, 21.7]	9.1 (6.9) [−4.4, 22.5]	15.3 (6.9) [1.8, 28.8]	−6.2 (9.7) [−25.3, 12.9]
Quotative *like*				
Scottish participants	72.3 (2.3) [67.8, 76.8]	84.4 (2.7) [79.1, 89.7]	60.2 (3.7) [53.1, 67.4]	24.1 (4.6) [15.2, 33.1]
English participants	52.6 (3.0) [46.6, 58.5]	67.9 (4.3) [59.5, 76.2]	37.3 (4.3) [28.9, 45.8]	30.5 (6.1) [18.6, 42.4]
Difference (Scot. – Eng.)	19.7 (3.8) [12.3, 27.2]	16.5 (5.1) [6.6, 26.5]	22.9 (5.7) [11.8, 34.0]	−6.4 (7.6) [−8.5, 21.3]

Figure 4 presents predictions based on the regression model, along with 95% confidence intervals for the estimates. The numeric values underlying Figure 4 are listed in Table 4, in the columns headed "Informal Speech" and "Semi-formal Writing". Averaging over speech and writing, we observe that Scottish participants report a higher overall usage rate (55.4%, 95% CI [49.8%, 61.1%]) compared to the English participants (19.5% [14.0%, 25.0%]). At 35.9 percentage points [28.1, 43.8], the difference between these variety averages is substantial, with the confidence interval suggesting that the higher overall currency in Scottish English is statistically dependable. For both varieties, we observe a stylistic cline that indicates that the sentence is more likely to be used in informal speech than in semi-formal writing. For the Scottish subgroup, the stylistic cline is 29.9 percentage points [18.6, 41.2], and the English respondents report a cline of 17.4 points [6.5, 28.3]. The difference in stylistic clines is 12.5 points [−3.2, 28.2] and estimated with a considerable margin of error. Nevertheless, there appears to be some indication for more pronounced register differences in the Scottish ratings.

Note that this (second) difference represents the interaction between Variety and Mode on the percentage scale.

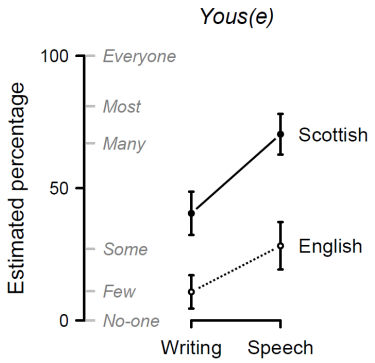


Figure 4. Model-based predictions for *yous(e)*; error bars denote 95% confidence intervals.

5.2. Extended use of the progressive

The data for the extended use of the progressive are summarized in Table 3 and Figure 5. Since two questionnaire items[4] targeted this eWAVE feature, the number of responses for each condition (i.e. variety-mode combination) is roughly twice the number of informants in the study. Figure 5 shows a steady progression from the written ratings by our English informants to the spoken ratings for our Scottish informants. This suggests that stative progressives, too, are generally (i.e., in both varieties) perceived to be less prevalent in writing, but they also appear to enjoy wider currency in the Scottish variety.

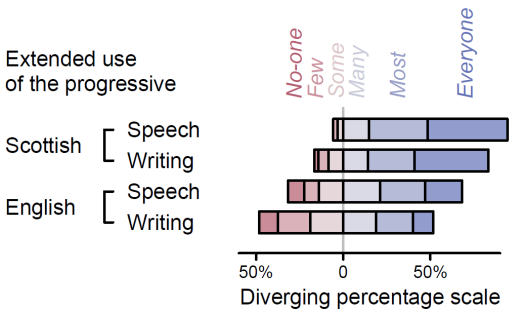


Figure 5. Distribution of ratings for the extended use of the progressive: Diverging bar chart.

The estimates from our regression model are displayed in Figure 6. Averaging over Mode, Scottish participants report a currency of 81.0% [77.8%, 84.1%], compared to 55.5% [50.8%, 60.2%] for the English subgroup. This translates into a difference of 25.5 percentage points [19.9, 31.2], which suggests that the data and model indicate a statistically reliable disparity between the varieties. The

[4] The two items are given as (2) and (3) in Section 3.2. Our analysis did include Item as a predictor, but for the purposes of the present paper, we report results averaged over its levels. Overall, however, the usage ratings were higher for *liking*, with a model-based prediction of 77% vs. 59% for *wanting* (averaging over Variety and Mode). Also see our discussion in Section 7.1.

Scottish subgroup reports a stylistic cline of 7.7 percentage points [1.4, 14.0], the English informants a cline of 12.5 points [3.1, 21.8]). The difference in stylistic clines therefore amounts to a modest −4.8 [−16.1, 6.5], which indicates that the standard vs. non-standard contrast is broadly similar in the two varieties.

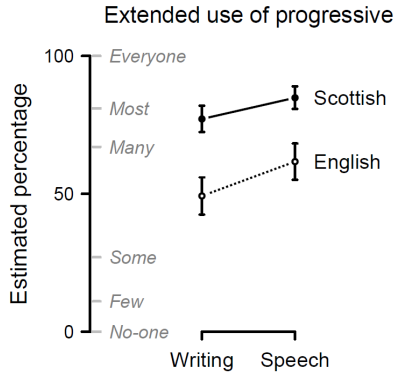


Figure 6. Model-based predictions for the extended use of the progressive; error bars denote 95% confidence intervals.

5.3. Epistemic *mustn't*

Epistemic *mustn't* is similarly common in both varieties. Figure 7 suggests that while usage ratings do not vary considerably across conditions, the Scottish informants report a somewhat higher occurrence for their variety.

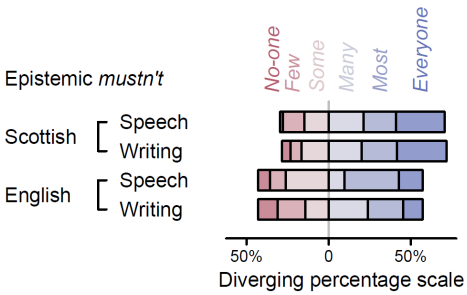


Figure 7. Distribution of ratings for epistemic *mustn't*: Diverging bar chart.

As is evident from Figure 8, the regression analysis confirms this trend: Usage reports average at 66.1% [60.2, 72.0] for Scottish English, compared to 53.9% [46.4, 61.4] for the English ratings. The observed difference of 12.2 percentage points [2.6, 21.7] is relatively small, but its 95% confidence interval suggests that the data and model are consistent with a higher currency in Scottish English. Interestingly, there is virtually no stylistic cline in the Scottish responses (−1.1 points [−12.9, 10.8]); according to the English participants, on the other hand, the feature is slightly more current in speech (a difference of 5.1 percentage points [−9.8, 20.1]), but the confidence interval shows that a broad range of differences (including negative ones) are plausible, given our data and model.

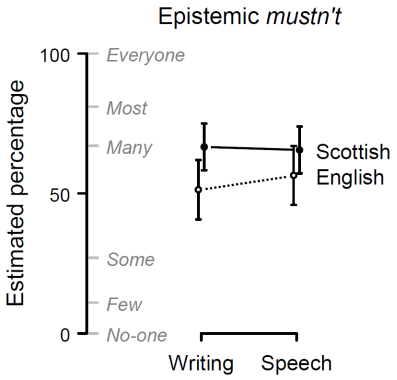


Figure 8. Model-based predictions for epistemic *mustn't*; error bars denote 95% confidence intervals.

5.4. Quotative *like*

The distribution of ratings for quotative *like* is presented in Table 3. The graphical arrangement in Figure 9 shows that both factors appear to play a noticeable role: A stylistic cline is clearly visible for both varieties, and the Scottish pair of bars extends further to the right.

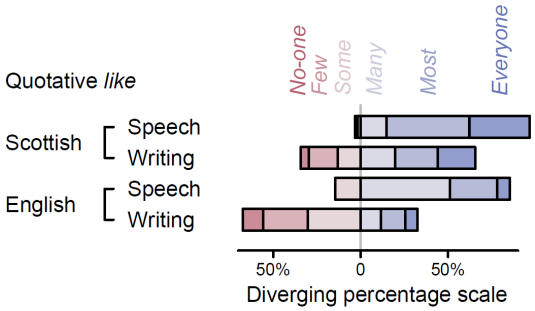


Figure 9. Distribution of ratings for quotative *like*: Diverging bar chart.

This also materializes in the regression results, which appear in Figure 10. Scottish ratings center on 72.3% [67.8%, 76.8%], and the English informants report an average usage rate of 52.6% [46.6%, 58.5%]. This yields an overall difference of 19.7 percentage points [12.3, 27.2], with the confidence interval giving us little reason to doubt that quotative *like* is more common in Scottish English. The stylistic clines are quite similar in the two varieties: The Scottish participants report a cline of 24.1 percentage points [15.2, 33.1], and the English responses translate into a cline of 30.5 points [18.6, 42.2].

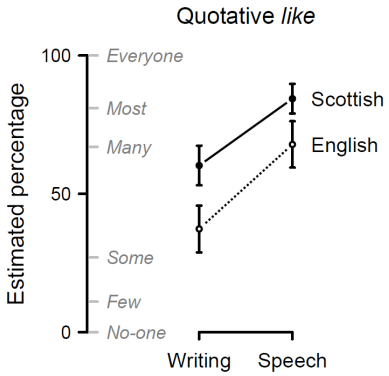


Figure 10. Model-based predictions for quotative *like*; error bars denote 95% confidence intervals.

6. Summary and discussion

Our questionnaire data indicate a higher currency of all four features in Standard Scottish English, which is consistent with the information we retrieved from eWAVE as well as the suggestions in the literature. We are therefore able to report empirical evidence that sets SSE apart from SSBE with regard to the morpho-syntactic features considered. However, we observe gradience among the features in terms of their function as variety markers: *Yous*(*e*) contrasts most clearly between the varieties (with an overall difference of 36 percentage points), followed by extended progressives (26 points) and quotative *like* (20 points); epistemic *mustn't*, on the other hand, shows a minor – though statistically reliable – difference (12 points).

It is also of interest to compare the estimated usage rates in SSE with the data in eWAVE. For a direct comparison of the sources, we refer to participant reports for the spoken setting. Indeed, we observe the highest currency estimates for the two features that receive A ratings in eWAVE ("feature is pervasive or obligatory"), namely the extended use of the progressive (85%) and quotative *like* (84%). The features with B ratings ("feature is neither pervasive nor extremely rare"), *yous*(*e*) and epistemic *mustn't*, are also judged as less prevalent by our Scottish informants, with ratings of 70% and 66%, respectively. This convergence strengthens our confidence in the BSLVC results, and it can also be considered as an external validation of the eWAVE report provided for Scottish English by Smith (2012).

Our data also allow us to assess the relative importance of varietal vis-à-vis stylistic differences. We observe that quotative *like* is the only feature where the stylistic stratification (an overall difference of 27 points, averaging over Variety) outweighs the difference between SSE and SSBE, which amounts to 20 points. For the other features, register differences are smaller than the varietal contrast. Nevertheless, *yous*(*e*) showed a pronounced overall stylistic cline of 24 points, followed by extended progressives (10 points) and epistemic *mustn't*, for which the overall level of stylistic variation is very low (2 points).

It is striking that in quotative *like* and *yous*(*e*), a pronounced cross-varietal awareness of the stylistic (non-standard) value of a feature coincides with a large inter-varietal difference. We would argue that such items may be particularly effective in indexing a variety, as they will be quite salient (in the non-technical sense of 'noticeable') for speakers of other varieties. If we consider them candidate features (or potential feature combinations) of SSE, they should perhaps be monitored particularly closely in future research, because some levelling – either of the difference to SSBE or of the stylistic value – might be expected.

In terms of stylistic clines, i.e. the extent to which a feature is reported to vary between speech and writing in Scottish English, we can identify two general pairs. On the one hand, quotative *like* and *yous*(*e*) respond very strongly to the stylistic dimension. This might be due to different reasons, since *like* in its quotative use originates in varieties of English outside Scotland (see the "Valley Girl" association mentioned in Section 3.4), whereas *yous*(*e*) is a classic feature of Scots.[5] In both cases, however, the stigma traditionally attached to the form may affect its present-day stylistic value. On the other hand, the use of stative verbs in the progressive does not respond very strongly, and epistemic *mustn't* hardly responds at all to style. In both cases, the difference from canonical standard grammar does not lie in the form of a construction but in its meaning. This may be the reason why they are less salient and accordingly less sensitive to prescriptivism and differences in stylistic context. As a point of departure for future research we would suggest that features of this type are promising candidates for the grammar of SSE.

7. Conclusion: BSLVC data in research on varieties of English

In the past two decades, corpora have become a (if not the) primary source of evidence in linguistics (see Palacios Matrínez 2020; Kortmann 2021), and this applies in particular to the World Englishes paradigm (see Lange and Leuckert 2020). Thus, of the 117 empirical papers published in the journal *English World-Wide* between 2012 and 2022, 56% relied on corpus data, the most prevalent resource being the family of ICE corpora (Greenbaum and Nelson 1996). Another resource that has been well-received in the community is the eWAVE atlas (Kortmann et al. 2020). As we saw in Section 3, this standard reference work is an invaluable source of information for a bird's-eye perspective on the regional distribution of a wide range of features. In this section, we discuss the status of BSLVC data as a source of empirical information vis-à-vis ICE corpora and eWAVE, which allows us to highlight some of its advantages and limitations.

7.1. Limitations

Let us begin with a consideration of the shortcomings of the BSLVC. Due to its breadth of coverage (128 features, of which 62 are directly informed by eWAVE), the survey includes only few items (in most cases only one) per morpho-syntactic structure. This severely limits the generality of findings derived from the questionnaire, and it carries the danger of other (unintended) characteristics of the sentence influencing – and thus potentially confounding – the responses given by informants. This means that extra caution must be exercised when interpreting findings derived from the BSLVC. Strictly speaking, differences between varieties, and between registers within a variety apply to the (set of) sentence(s) presented to the informants. Corpora and eWAVE do not have this drawback: With the exception of infrequent structures, corpora usually provide a more solid empirical basis and include multiple occurrences of a feature in natural language use. Similarly, the expert judgments in eWAVE apply to spoken language in general, and these assessments therefore naturally need to cover a wide range of styles.

Further, while the BSLVC elicits ratings for two registers (informal speech and semi-formal writing), the very limited number of linguistic contexts (i.e. sentences) presented to informants precludes the study of internal variables and lexical effects in the usage pattern of a given structure. For instance, the BSLVC provides no insight into which stative verbs other than LIKE and WANT tend to be used

[5] There is a word-formation process in Scots, whereby the affix *-like* modifies the meaning of an adjective or at least generates a formal variant, as in *wise-like* 'presentable' or *daft-like* 'silly' (Grant & Dixon 1921: 185), but we do not attempt to link this to present-day usage rates of the quotative construction.

with the progressive, and the question of whether quotative *like* primarily serves to introduce dramatized personal narrative or whether it has developed into a neutral quotative device remains open. While this shortcoming also applies to eWAVE, corpora do offer the opportunity to study regularities underlying the patterned heterogeneity that is commonly observed in language variation and change.

A further limitation of the BSLVC is the fact that data for the grammar part are obtained from university and secondary school students. While this is mainly due to practical considerations (see Krug and Sell 2013: 79-84), for some varieties (but of course not for Scotland), this may be the only sub-population that actually uses English to a considerable extent. The BSLVC shares this drawback with the ICE family corpora, which likewise target educated speakers.

Finally, since many of the linguistic phenomena covered by the BSLVC are colloquial, dialectal, or non-standard – and therefore stigmatized among the highly educated strata of society – social desirability effects may affect the judgments reported by individuals. While this distorting effect cannot be dismissed entirely, the BSLVC survey tries to circumvent this issue by asking informants about the language use of the broader speech community (rather than their own use, or that of their friends/relatives).

7.2. Advantages

An important asset of the BSLVC is the level of comparability and control it offers: All participants fill in the same survey, which therefore yields a high level of comparability between socio-demographic subgroups and varieties. Its standardized form is an advantage compared to ICE corpora, whose direct comparability has been the subject of some discussion (e.g. Vetter 2021), and also the eWAVE database, where each variety is profiled by a different expert (see Kortmann and Lunkenheimer 2012: 6 for a discussion of limitations).

A further attractive aspect of the BSLVC is its elicitation of judgments for two registers, informal speech and semi-formal writing. This is in contrast to eWAVE, which provides ratings for spoken language only. As we have seen in the analysis reported in the present paper, the ability to differentiate two styles and to quantify the stylistic cline associated with a particular feature holds considerable research potential. For instance, quotative *like* differs from the other features in that Mode accounts for a greater difference in usage rate than Variety. From a statistical perspective, the fact that participants provide judgments for both styles offers an advantage: Style becomes a within-subject factor and is therefore (usually) estimated with higher level of precision.

Questionnaires and surveys are also attractive for the study of infrequent features, for which ICE corpora only yield sparse data (e.g. epistemic *mustn't*). Further, the usage rate of certain structures can only meaningfully be expressed by reference to choice contexts (see Wallis and Mehl 2022), or an appropriate envelope of variation. This may prove to be rather challenging (e.g. for the progressive, but see Rautionaho 2022) or laborious (e.g. for quotative *like*, which requires a consideration of all quotative devices).

Another asset of the BSLVC is its rich sociodemographic metadata, which allows for an exploration of the external dimension of language variation. It should be noted, however, that the grammar part is mostly administered to university students, which leads to a rather homogeneous sample (young and educated speakers). Closely matching samples of the population in different varieties, of course, entail both advantages (such as comparability) as well as disadvantages. Differences in regional background, however, may offer clues to between-speaker variability. For *yous*(*e*), for instance, a

natural line of further inquiry is whether there is a north-south cline among the English informants, which would be expected on the basis of our literature review (see Table 2).

A further distinction of the BSLVC compared to eWAVE is that it provides crowd-sourced estimates of the prevalence of structures in a variety. Such reports from the coalface of language use are valuable in that they may corroborate or qualify the ratings provided by experts. And the analyses presented in the present study show that the BSLVC has the potential to offer currency ratings at a finer level of detail than eWAVE.

Overall, then, we have seen how the questionnaire-based insights provided by the BSLVC can enrich the empirical landscape for research on World Englishes. Its advantages and drawbacks tie in well with existing and widely used resources such as the ICE corpora and eWAVE: Its weaknesses are to a certain extent cushioned by these alternatives, and its strengths allow it to make genuine contributions to the documentation of spoken and written varieties. As a complementary line of empirical evidence, then, the BSLVC offers opportunities for methodological triangulation and has the potential to add unique insights to our understanding of Englishes world-wide.

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. Chambers.

Anderson, Wendy & John Corbett. 2008. The Scottish Corpus of Texts and Speech – a user's guide. *Scottish Language* 27. 19–41.

Anderwald, Lieselotte. 2020. English dialects in the Southeast of England. In: Kortmann, Bernd & Lunkenheimer, Kerstin & Ehret, Katharina (eds.), *The Electronic World Atlas of Varieties of English*. http://ewave-atlas.org/languages/9 (accessed on 2024-09-12).

Arel-Bundock, Vincent. 2023. *marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests*. R package version 0.13.0, https://CRAN.R-roject.org/package=marginaleffects.

Beal, Joan C. 1993. The grammar of Tyneside and Northumbrian English. In James Milroy & Lesley Milroy (eds.), *Real English: The grammar of English dialects in the British Isles*, 187–242. Longman.

Beal, Joan C. 1997. Syntax and morphology. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, 335–377. Edinburgh University Press.

Beal, Joan C. 2010. An introduction to regional Englishes: Dialect variation in England. Edinburgh University Press.

Blyth, Carl, Sigrid Recktenwald & Jenny Wang. 1990. I'm like, 'Say what?!': A new quotative in American oral narrative. *American Speech* 65(3). 215–227.

Britain, David. 2020. Grammatical variation in the contemporary spoken English of England. In A. Kirkpatrick (ed.), *The Routledge handbook of World Englishes*, 32–58. Routledge.

Buchstaller, Isabelle. 2006. Diagnostics of age-graded linguistic behaviour: The case of the quotative system. *Journal of Sociolinguistics* 10. 3–30.

Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of quotative 'be like'. *Journal of Sociolinguistics* 13(3). 291–331.

Butters, Ronald R. 1980. Narrative go 'say'. *American Speech* 55(4). 304–307.

Cheshire, Jenny, Paul Kerswill, Susan Fox & Eivind N. Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics* 15. 151–196.

Corbett, John, J. Derrick McClure &Jane Stuart-Smith (eds.) 2003. *The Edinburgh companion to Scots*. Edinburgh University Press.

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 University Press.
Crawford, Naomi. 2017. Do people tweet like they speak? A study of intraspeaker variation. *Lifespans and Styles* 3(2). 29–39.
Douglas, Fiona M. 2003. The Scottish Corpus of Texts and Speech: Problems of corpus design. *Literary and Linguistic Computing* 18(1). 23–37.
Durham, Mercedes, Bill Haddican, Eytan Zweig, Daniel E. Johnson, Zipporah Baker, David Cockeram, Esther Danks & Lousie Tyler. 2012. Constant linguistic effects in the diffusion of be like. *Journal of English Linguistics* 40(4). 316–337.
Ferrara, Kathleen & Barbara Bell. 1995. Sociolinguistic variation and discourse function of constructed dialogue introducers: The case of *be + like*. *American Speech 70*. 265–290.
Filppula, Markku. 2020. Irish English. In Kortmann, Bernd & Kerstin Lunkenheimer & Katharina Ehret (eds.), *The Electronic World Atlas of Varieties of English*. http://ewave-atlas.org/languages/3 (accessed on 2024-09-12).
Gardner, Matt H., Derek Denis, Marisa Brook & Sali Tagliamonte. 2021. Be like and the Constant Rate Effect: From the bottom to the top of the S-curve. *English Language and Linguistics* 25(2). 281–324.
Giegerich, Heinz J. 1992. *English Phonology*. Cambridge University Press.
Grant, William & J. Main Dixon. 1921. *Manual of Modern Scots*. Cambridge University Press.
Greenbaum, Sidney & Gerald Nelson. 1996. The International Corpus of English (ICE) Project. *World Englishes* 15(1). 3–15.
Heiberger, Richard M. & Naomi B. Robbins. 2014. Design of diverging stacked bar charts for Likert scales and other applications. *Journal of Statistical Software* 57(5). 1–32.
Knapp, Thomas R. 1990. Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research* 39(2). 121–123.
Kortmann, Bernd. 2008. Synopsis: Morphological and syntactic variation in the British Isles. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English 1: The British Isles*, 478–495. Mouton de Gruyter.
Kortmann, Bernd. 2021. Reflecting on the quantitative turn in linguistics. *Linguistics* 59(5). 1207–1226.
Kortmann, Bernd & Kerstin Lunkenheimer (eds.). 2012. *The Mouton world atlas of variation in English*. Walter de Gruyter.
Kortmann, Bernd, Kerstin Lunkenheimer & Katharina Ehret. 2020. *The Electronic World Atlas of Varieties of English*. https://doi.org/10.5281/ZENODO.3712132.
Kranich, Svenja. 2010. *Progressive in Modern English: A corpus-based study of grammaticalization and related changes*. Rodopi.
Krug, Manfred & Katrin Sell. 2013. Designing and conducting interviews and questionnaires. In Manfred Krug & Julia Schlüter (eds.), *Research methods in language variation and change*, 69–98. Cambridge University Press.
Krug, Manfred, Ole Schützler, Fabian Vetter & Lukas Sönning. 2025. Background data for: The morpho-syntax of Scottish Standard English: Questionnaire-based insights. https://doi.org/10.18710/B3NJBT, DataverseNO, V1.
Lange, Claudia & Sven Leuckert (eds.). 2020. *Corpus linguistics for World Englishes: A guide for research*. Taylor and Francis.
Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge University Press.
Levin, Magnus. 2013. The progressive verb in modern American English. In Bas Aarts, Joanne Close, Geoffrey Leech & Sean Wallis (eds.), *The verb phrase in English: Investigating recent language change with corpora*, 187–216. Cambridge University Press.

Macaulay, Ronald K. S. 2005. *Talk that counts: Age, gender, and social class differences in discourse.* Oxford University Press.
MacKenzie, Laurel, George Bailey & Danielle Turton. 2022. Towards an updated dialect atlas of British English. *Journal of Linguistic Geography* 10(1). 46–66.
Maguire, Warren. 2012. English and Scots in Scotland. In Raymond Hickey (ed.), *Areal features of the anglophone world,* 53–77. De Gruyter Mouton.
Macafee, Caroline & Colm Ó Baoill. 1997. Why Scots is not a Celtic language. In Hildegard L. C. Tristram (ed.), *The Celtic Englishes*, 245–286. Winter.
McArthur, Tom. 1987. The English languages. *English Today* 3(3). 9–13.
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 University Press
Meyerhoff, Miriam & Erik Schleef. 2014. Hitting an Edinburgh target: Immigrant adolescents' acquisition of variation in Edinburgh English. In Robert Lawson (ed.), *Sociolinguistics in Scotland*, 103–128. Palgrave Macmillan.
Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics, 15*(1), 1–67.
Miller, Jim. 1993. The grammar of Scottish English syntax. In James Milroy and Lesley Milroy (eds), *Real English: The grammar of English dialects in the British Isles*, 99–138. Longman.
Miller, Jim. 2008. Scottish English: Morphology and Syntax. In Bernd Kortmann & Clive Upton (eds.), *Varieties of English 1: The British Isles*, 299–327. Mouton de Gruyter.
Newstead, S. E., P. Pollard & D. Riezebos. 1987. The effect of set size on the interpretation of quantifiers used in rating scales. *Applied Ergonomics* 18(3). 178–182.
Palacios Martínez, Ignacio M. 2020. Methods of data collection in English empirical linguistics research: Results of a recent survey. *Language Sciences* 78. 101263.
Papke, Leslie E. & Jeffrey M. Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics* 11(6). 619-632.
Pearce, Michael. 2021. The participatory vernacular web and regional dialect grammar: A new account of pronouns in North East England. *English Today* 37(4), 196–205.
Penhallurick, Robert. 2020. Welsh English. In: Kortmann, Bernd & Lunkenheimer, Kerstin & Ehret, Katharina (eds.), *The Electronic World Atlas of Varieties of English.* http://ewave-atlas.org/languages/5 (accessed on 2024-09-12).
R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rautionaho, Paula. 2022. Revisiting the myth of stative progressives in world Englishes. *World Englishes* 41. 183–206.
Rautionaho, Paula & Robert Fuchs. 2021. Recent change in stative progressives: a collostructional investigation of British English in 1994 and 2014. *English Language and Linguistics*, 25(1). 35–60.
Richards, Hazel M. 2008. Mechanisms, motivations and outcomes of change in Morley (Leeds) English. University of York dissertation.
Sarkar, Deepayan. 2008. *Lattice: Multivariate data visualization with R.* Springer.
Schützler, Ole. 2024. The elusive butterfly of Scottish Standard English. In Christine Elsweiler (ed.), *The languages of Scotland and Ulster in a global context, past and present. Selected Papers from the 13th triennial Forum for Research on the Languages of Scotland and Ulster, Munich 2021.* FRLSU. 91–138.
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. Mouton de Gruyter.

Schützler, Ole & Jenny Herzky. 2021. Modal verbs of strong obligation in Scottish Standard English. *English Language and Linguistics* 26(1). 133–159.
Smith, Jennifer. 2012. Scottish English and varieties of Scots. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 21–29. Walter de Gruyter.
Smith, Jennifer. 2020. Scottish English. In Kortmann, Bernd & Kerstin Lunkenheimer & Katharina Ehret (eds.), *The Electronic World Atlas of Varieties of English*. http://ewave-atlas.org/languages/2 (accessed on 2024-09-12).
Sönning, Lukas. 2024. Ordinal response scales: Psychometric grounding for design and analysis. *Research Methods in Applied Linguistics* 3(3). 100156.
Stenström, Anna-Brita, Gisle Andersen & Ingrid K. Hasund. 2002. *Trends in teenage talk: Corpus compilation, analysis and findings*. Benjamins.
Stevens, S.S. 1946. On the theory of scales of measurement. *Science* 103 (2684). 677–680.
Svartvik, Jan & Geoffrey Leech. 2016. *English: One tongue, many voices*. Palgrave Macmillan.
Tagliamonte, Sali & Rachel Hudson. 1999. Be like et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3. 147–172.
Trousdale, Graeme. 2020. English dialects in the North of England. In Kortmann, Bernd & Kerstin Lunkenheimer & Katharina Ehret (eds.), *The Electronic World Atlas of Varieties of English*. http://ewave-atlas.org/languages/6 (accessed on 2024-09-12).
Vetter, Fabian. 2021. *Issues of corpus comparability and register variation in the International Corpus of English: Theories and computer applications*. University of Bamberg dissertation.
Wagner, Susanne. 2020. English dialects in the Southwest of England. In Kortmann, Bernd & Kerstin Lunkenheimer & Katharina Ehret (eds.), *The Electronic World Atlas of Varieties of English*. http://ewave-atlas.org/languages/7 (accessed on 2024-09-12).
Wallis, Sean & Seth Mehl. 2022. Comparing baselines for corpus analysis: Research into the *get*-passive in speech and writing. In Ole Schützler & Julia Schlüter (eds.), *Data and methods in corpus linguistics: Comparative approaches*. 101–126. Cambridge University Press.
Wells, John C. 1982. *Accents of English*. 3 vols. Cambridge University Press.